

Nuclear Heat Transfer and Passive Cooling

Volume 4: Confidence and Uncertainty



Introduction to the Technical Volumes and Case Studies



Convection, Radiation and Conjugate Heat Transfer



Natural Convection and Passive Cooling



Confidence and Uncertainty



Liquid Metal Thermal Hydraulics



Molten Salt Thermal Hydraulics



Liquid Metal CFD Modelling of the TALL-3D Test Facility



Fuel Assembly CFD and UQ for a Molten Salt Reactor



Reactor Scale CFD for Decay Heat Removal in a Lead-cooled Fast Reactor



System Code and CFD Analysis for a Light Water Small Modular Reactor

Authors:	Graham Macpherson	Frazer-Nash Consultancy
	Neil Headings	Frazer-Nash Consultancy
	Ruth Aldred	Frazer-Nash Consultancy
Contributors:	Carolyn Howlett	Frazer-Nash Consultancy
	Adam Owens	Moltex Energy
	Milorad Dzodzo	Westinghouse Electric Company
	Gregory Banyay	Westinghouse Electric Company and Pennsylvania State University
	Steven Moor	Rolls-Royce
Technical Volumes Lead:	Tim Houghton	Frazer-Nash Consultancy
Approver:	Richard Underhill	Frazer-Nash Consultancy
Document Number:	FNC 60148/49311R	
Issue and Date:	Issue 1, December 2021 (minor typographical update, March 2023)	

Legal Statement

This document presents work undertaken by Frazer-Nash Consultancy Ltd and funded under contract by the UK Government Department for Business, Energy and Industrial Strategy (BEIS). Any statements contained in this document apply to Frazer-Nash Consultancy and do not represent the views or policies of BEIS or the UK Government. Any copies of this document (in part or in full) may only be reproduced in accordance with the below licence and must be accompanied by this disclaimer.

This document is provided for general information only. It is not intended to amount to advice or suggestions on which any party should, or can, rely. You must obtain professional or specialist advice before taking or refraining from taking any action on the basis of the content of this document.

We make no representations and give no warranties or guarantees, whether express or implied, that the content of this document is accurate, complete, up to date, free from any third party encumbrances or fit for any particular purpose. We disclaim to the maximum extent permissible and accept no responsibility for the consequences of this document being relied upon by you, any other party or parties, or being used for any purpose, or containing any error or omission.

Except for death or personal injury caused by our negligence or any other liability which may not be excluded by an applicable law, we will not be liable to any party placing any form of reliance on the document for any loss or damage, whether in contract, tort (including negligence) breach of statutory duty, or otherwise, even if foreseeable, arising under or in connection with use of or reliance on any content of this document in whole or in part.

Unless stated otherwise, this material is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. You may copy and redistribute the material in any medium or format, provided you give appropriate credit, provide a link to the license and indicate if changes were made. If you remix, transform, or build upon the material, you may not distribute the modified material. You may not restrict others from doing anything the license permits.



Preface

Nuclear thermal hydraulics is the application of thermofluid mechanics within the nuclear industry. Thermal hydraulic analysis is an important tool in addressing the global challenge to reduce the cost of advanced nuclear technologies. An improved predictive capability and understanding supports the development, optimisation and safety substantiation of nuclear power plants.

This document is part of *Nuclear Heat Transfer and Passive Cooling: Technical Volumes and Case Studies*, a set of six technical volumes and four case studies providing information and guidance on aspects of nuclear thermal hydraulic analysis. This document set has been delivered by Frazer-Nash Consultancy, with support from a number of academic and industrial partners, as part of the UK Government *Nuclear Innovation Programme: Advanced Reactor Design*, funded by the Department for Business, Energy and Industrial Strategy (BEIS).

Each technical volume outlines the technical challenges, latest analysis methods and future direction for a specific area of nuclear thermal hydraulics. The case studies illustrate the use of a subset of these methods in representative nuclear industry examples. The document set is designed for technical users with some prior knowledge of thermofluid mechanics, who wish to know more about nuclear thermal hydraulics.

The work promotes a consistent methodology for thermal hydraulic analysis of single-phase heat transfer and passive cooling, to inform the link between academic research and end-user needs, and to provide a high-quality, peer-reviewed document set suitable for use across the nuclear industry.

The document set is not intended to be exhaustive or provide a set of standard engineering 'guidelines' and it is strongly recommended that nuclear thermal hydraulic analyses are undertaken by Suitably Qualified and Experienced Personnel.

The first edition of this document set has been authored by Frazer-Nash Consultancy, with the support of the individuals and organisations noted in each. Please acknowledge these documents in any work where they are used:

Frazer-Nash Consultancy (2021) Nuclear Heat Transfer and Passive Cooling,
Volume 4: Confidence and Uncertainty.

Contents

1	Introduction and Technical Context	1
1.1	Building Confidence in Nuclear Thermal Hydraulic Models	2
1.1.1	Model Design and Construction	3
1.1.2	Verification and Quality Assurance	5
1.1.3	Validation	6
1.1.4	Uncertainty Quantification and Sensitivity Analysis	7
1.1.5	Overall Estimate of Confidence in a Prediction	8
1.2	Uncertainty	8
1.3	Thermal Hydraulic Phenomena that Contribute Uncertainty	11
1.4	Why are V&V and UQ Hard for Industrial Users?	14
2	Concepts for Uncertainty	16
2.1	Sources and Treatment of Uncertainty	16
2.1.1	Common Sources of Uncertainty	16
2.1.2	Treatment of Different Types of Uncertainty	20
2.1.3	Significance of Output Metrics	20
2.1.4	Initial Sensitivity Assessment	22
2.2	Common Methods and Tools	23
2.2.1	Basic Statistics	23
2.2.2	Toolkits	23
2.2.3	Visualisation	24
2.3	Propagation of Uncertainty	26
3	Methodologies	28
3.1	Organisational Approaches	28
3.1.1	EMDAP, CSAU and PIRT	28
3.1.2	Predictive Capability Maturity Model	29
3.1.3	Approaches from Other Industries	30
3.1.4	Examples of PIRT and PCMM	30
3.2	Approaches to Component Activities	31
3.2.1	Frameworks and Guidance for V&V and UQ	33
3.2.2	Obtaining Validation Data	34
3.2.3	Making Comparisons Between Validation Data and Simulations	35
3.2.4	How and When to Use UQ and SA	37
3.3	Methods for System Codes	38
3.4	Methods for CFD	40
3.4.1	Model and Numerical Uncertainty in CFD	41
3.4.2	Validation Data and Benchmarks	42
3.4.3	Applying Polynomial Chaos	45
3.4.4	Proper Orthogonal Decomposition and Machine Learning	46

3.5	Applying Effort in the Right Place	47
3.5.1	Change Control and the Validity of Geometry and Input	47
3.5.2	Code and Solution Verification vs. Practical CFD Models	48
3.5.3	Should UQ be Applied to Every Analysis?	50
4	Mathematical Methods for Propagating Uncertainty	51
4.1	Conceptual Statistical Framework – Frequentist vs Bayesian	51
4.2	Parametric Uncertainty Characterisation	52
4.2.1	Considerations When Characterising an Uncertain Input	52
4.2.2	Example of Considerations for an Uncertain Input	55
4.2.3	Central Limit Theorem	56
4.2.4	Characterising Uncertain Inputs	56
4.3	Sampling or Design of Computer Experiments	67
4.3.1	Structured Grid Based Methods	68
4.3.2	Unstructured Methods	70
4.3.3	Selecting a Method	72
4.4	Surrogate Models	73
4.4.1	Requirements of a Surrogate Model	74
4.4.2	Types of Surrogate Model	75
4.5	Sensitivity Analysis	80
4.5.1	Scatter Plots	81
4.5.2	Sensitivity Coefficients	81
4.5.3	Correlation Coefficients	81
4.5.4	Sobol Indices	82
4.6	Uncertainty Quantification by Propagation	83
4.6.1	Use of Surrogate Models	83
4.6.2	Reliability Methods	84
4.7	Assessment of Output Uncertainty from Propagation	85
4.7.1	Confidence Intervals and Levels	85
4.7.2	Statistical Tolerance Limits	87
4.7.3	Chebyshev's Rule	88
4.8	Combining Uncertainty	88
4.8.1	Propagating Epistemic and Aleatory Uncertainties	88
4.8.2	Model Uncertainty	89
4.8.3	Impact on the Number of Model Calculations Required	91
4.8.4	Inter-model Comparison	91
4.8.5	Presenting Overall Uncertainty in an Output Metric	92
5	References	94
6	Abbreviations	106

1 Introduction and Technical Context

A necessary aspect of nuclear engineering professionalism includes having a questioning attitude. In the context of numerical models used to provide predictions across all aspects modelling and simulation, this is realised by not simply accepting or assuming that the application of, and outputs from, a software tool are correct without questioning a number of aspects. These are relevant even for widely accepted and applied tools:

1. How will the results be used?
2. Is the simulation 'right' and how much can I trust it – how adequate and realistic are the inputs and results?
3. Am I using the tool within its design or its intended and trusted prediction envelope?
4. Have I made any errors, and how can I check this? What extent of independent review and checking is appropriate?
5. How can I convince others, who will need to use the information to make a potentially costly or safety critical decision, of the correctness of the results? Can I clearly communicate the extent of, or reasons for, my level of confidence?
6. How can I help others understand and use these results, whilst ensuring that they are not used out of context or inappropriately? Particularly because end users are unlikely to have the same specialised domain knowledge and case-specific experience as the originator of the results.
7. How much effort is warranted in considering these aspects, given the intended use of the results?

Pragmatism will be necessary when performing a modelling analysis and forming answers to these questions for an industrial purpose. It will usually not be possible to wait until there are sufficiently large computational resources, improved models and tools, or high quality experimental data to fully resolve all aspects (or to justify their expense) because the analysis results will be required to make a decision in a timely manner.

Alternatively, the decision being made may be sufficiently important that a substantial investment in extensive computing and experimental assessment proves to be necessary. This could cause increased costs and operational or construction delays. Therefore, the evidence and justification for extensive modelling activities relies on knowing, with enough rigour and precision, which aspects are most uncertain, which aspects matter most, and what extent of further evaluation or assessment is required.

Any assessment of Nuclear Thermal Hydraulics (NTH) model calculations with a nuclear safety implication must be aligned with the expectations of a regulator (Volume 1, Section 2.2). In particular,

Introduction and Technical Context

the application of a graded approach, where the effort and rigour should be proportionate to the significance of the decision that an analysis supports.

This technical volume aims to provide assistance by describing two interrelated themes, aimed at assessing the credibility that can be attributed to the results of NTH models:

Confidence: Good modelling and working practices to establish and communicate confidence¹ in thermal hydraulic models and their simulation results. This includes applying quality assurance and review practices, performing Verification and Validation (V&V), applying recognised best practices, documenting work, and archiving to allow its easy retrieval.

Uncertainty: Uncertainty represents an inability to state a definitively accurate and precise result for a modelling prediction. Here it involves understanding the types, sources and magnitudes of uncertainty in a thermal hydraulic analysis. This enables Uncertainty Quantification (UQ) to be applied to metrics of interest for reactor safety or performance to determine the significance of inherent variability, gaps in knowledge, or approximations and simplifications in a model. It also allows the evaluation of which model inputs or parameters matter most to a prediction via Sensitivity Analysis (SA).

V&V, UQ and SA are not specifically thermal hydraulic or nuclear industry techniques – they are large and established disciplines that are applied in many branches of engineering and science. Their relevance to NTH, and the scrutiny that they are subject to, is growing, however. Large increases in the availability of computational resources and advanced methods such as multiphysics coupling have enabled increases in model complexity and scope. This has resulted in the wider application and greater importance of computational methods within design and safety.

The concepts that are needed for UQ² are introduced in Section 2. Combining these approaches and methods and applying them to practical industrial NTH analysis using system codes and Computational Fluid Dynamics (CFD) models is described in Section 3, then mathematical methods that are applied for UQ are detailed in Section 4.

1.1 Building Confidence in Nuclear Thermal Hydraulic Models

Using model predictions with confidence requires individuals and organisations to not only perform specific mathematical and computational tasks, but to understand the reasons why they are performing them, and to fit their outcomes into a broader structure. A hierarchy of considerations is typically involved in establishing and maintaining this confidence, each tier being a subset and enabler of the tier above:

1. Organisational approaches that determine, control, record and communicate what activities are needed to support the application of analysis tools.
2. Activities that are component parts supporting an analysis performed within or contributing to the overarching organisational approach.
3. Mathematical methods and software tools that are applied in performing these activities.

¹ Here the word 'confidence' is used in a wide and qualitative sense; it also has a specific quantitative meaning in relation to confidence intervals, as discussed in Section 4.7

² When UQ is discussed in this technical volume, the concepts and methods are generally applicable to SA too, and although they are distinct activities, UQ will often be used for brevity to implicitly refer to both.

Introduction and Technical Context

These considerations correspond to Figure 1.1, which is intended to show the generic stages where decisions and evaluations are made when embarking on an analysis task for a particular purpose, and how they sit within a broader context. An analysis task has a set of requirements that define and initiate it, and its output will usually deliver a report that can inform a decision.

An analysis will sit within the organisational context above it (Tier 1 in the list above) and will perform component activities (from 'Model Design' to 'Overall Confidence', Tier 2) making use of methods and tools below (Tier 3).

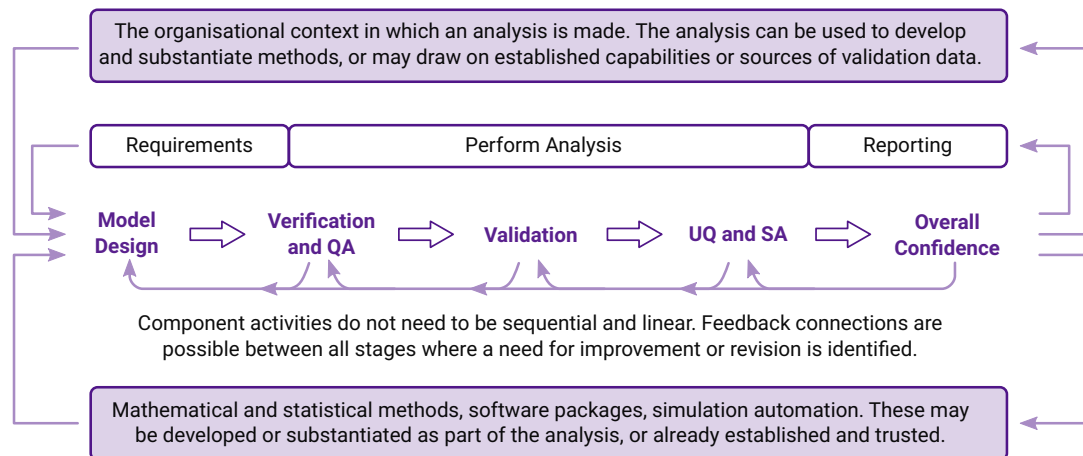


Figure 1.1: Activities to be performed to support having confidence in the results of an analysis task, and the context in which they take place.

The organisational context includes, for example, established Quality Assurance (QA) and V&V procedures and guidance that an organisation has defined to enable analysis methods to be applied in a consistent manner. The particular analysis being performed may be to solve a specific design or operational problem, or may be part of programme of developing the organisational approach. It may also have the purpose of specifically developing or enhancing a piece of numerical simulation software ('codes') or to produce solution automation methods. Alternatively, these may just arise as a valuable by-product, available for use in future analyses.

1.1.1 Model Design and Construction

The type of analysis that will be performed, which tools are used, and what type of activities are necessary to demonstrate confidence in the outputs depend on its intended use, as defined by its requirements³. A substantial number of decisions need to be made in choosing an overall approach and designing and constructing an NTH model. These decisions can significantly affect its eventual utility and cost:

- What is the safety or economic risk or importance influenced by how the output will be used? Does it inform concept design iterations, or is it evidence for a key safety case claim?
- What are the quantities of interest?

³ Systems engineering, as used to capture and demonstrate compliance with safety and operational requirements is an established discipline in itself, with its own methods and tools. Similarly, the creation of safety cases, using methods like Probabilistic Safety Analysis (PSA) is its own discipline, and is briefly discussed in Volume 1, Section 2.2. The requirements for an analysis are likely to arise from these activities, but details of this process are not covered in this volume.

Introduction and Technical Context

- Do the quantities of interest predicted by the model represent the mean behaviour of a system, or are they the extent (or chance) of the occurrence of an extreme aspect of its behaviour?⁴
- Is the method a standard technique or a bespoke and novel model? Is the use of an existing technique within the scope of its accepted purpose, or is it being used in a new context?
- What amount of extrapolation to 'new' areas of the parameter space is required? How far are new simulations outside the range of previous applications of the model, its established validation envelope, or the validity range of its assumptions and submodels?
- How computationally expensive is the model to run?
- How robust is the solution process?
 - Will it always either reliably converge, or fail to produce a solution in an informative and deterministic way?
 - Alternatively, is it prone to unpredictable numerical divergence, and/or can fail to run properly in a non-obvious way?

The latter type may be hard to avoid for complex models, and require significant manual inspection of the solutions and case-by-case intervention in their setup and solution process, limiting the amount of automation that can be applied to them.

- What existing guidance and best practice is available for this type of analysis? Has a search of the available literature been conducted?
- Are there regulatory or contractual requirements that mean that some methods must (or may not) be used?
- Who is performing the analysis and review? How experienced are the personnel? Do they have sufficient mentoring and supervision?

When using numerical models (NTH simulations being one example of many) to demonstrate confidence that a Nuclear Power Plant (NPP) will remain within safety limits, two categories of approach are generally considered.

Conservative approach: Historically, incomplete knowledge of plant behaviour and limited computing capability for modelling thermal hydraulic phenomena led to the application of a conservative (and thus bounding) methodology. This used both a conservative (or 'pessimistic') model/code and conservative values of model parameters to demonstrate a high level of confidence that a prediction bounded the worst-case (most onerous) behaviour of a system.

BEPU approach: More recently a Best Estimate Plus Uncertainty (BEPU) approach has increasingly been adopted. This uses a more realistic model of thermal hydraulic phenomena and more realistic input data with due allowance for uncertainties accounted for when analysing the model results.

The chosen approach can have a large disparity in cost and complexity, and both have a role to play at various stages of design or licensing, as determined by using a graded approach (Volume 1, Section 2.2). In both cases, steps are necessary to ensure that confidence in the thermal hydraulic

⁴ Another interpretation of this is to ask whether it is the combined worst cases of its aleatory uncertainty that are important? (Section 1.2)

Introduction and Technical Context

model is justified and appropriately communicated, and that adequate margin will be maintained to the safety limits.

The distinction of approaches into conservative and BEPU are concepts primarily developed for numerical modelling to support safety assessments. However, modelling is also performed for a range of design and accident scenario assessments that have a bearing on the operation and economic performance of a plant, and the decisions made based on it can have consequences for the viability of a technology or development project. Therefore, there are related motivations to know how much confidence to place in an assessment, and what steps will be most effective to increase that confidence.

1.1.2 Verification and Quality Assurance

Verification and quality assurance activities are grouped into three main categories.

Code Verification provides confidence that the mathematical equations have been implemented correctly and that the algorithms used to solve them behave as required in the software tool used.

Solution Verification of analyses is often defined as specifically assessing numerical error in specific computational simulations (rather than generically of the tool itself) arising mainly from iterative convergence and discretisation errors. In this context, it typically focuses primarily on how solution prediction converges as a result of increasing refinement of geometrical discretisation (mesh or nodalisation).

Quality Assurance Verification provides confidence that the model inputs and geometry have been correctly set up, that the model has been correctly solved, and that the results are extracted, analysed and presented properly. Some of these activities are included under descriptions of 'Solution Verification' in some references, but are often only given brief coverage. This ensures, through peer-review and independent inspection, code-to-code comparison, checking and approval that the model (including the software via specific Software Quality Assurance (SQA)) and its reporting do not contain errors, that they conform to specified standards and requirements, and that traceable evidence of this is produced.

Deciding which of these processes to apply, to what extent and by which methods is affected by several considerations:

- What procedures for verification and QA does the organisation instructing the analysis or regulator require to be followed, and what standard of record keeping and evidence is required?
- What are the key things to check – where are errors most likely to occur and which are of the most consequence?
- Is the decision relying on the result so significant that an independent (blind) parallel calculation need to be performed by a different group?
- Is additional code verification required, or is the code sufficiently mature?
- Could an analyst be expected to be knowledgeable enough about how a code is built, and how its numerics operate to meaningfully assess the correctness of its operation, or does it need one of the code developers to participate?

Introduction and Technical Context

- Are there sufficiently experienced analysts, who have been independent of the work, available to perform quality assurance verification?
- What are the resource requirements for these activities, and are their availability, time and costs allowed for? Verification and QA can represent an additional cost that is a significant proportion of the cost of originating an analysis and reporting it. Code developers or sufficiently experienced verifiers and reviewers also may only exist outside of the organisation performing the analysis.

1.1.3 Validation

Validation has a number of generally agreed and compatible definitions⁵ – in this technical volume it is conceptualised as:

Validation: Comparison of model predictions against measurements, to provide confidence that simulation results represent reality with sufficient accuracy.

There are questions that should be considered related to the availability, quality of and application of experimental or plant data:

- What data is available to validate a model against?
- Does the data come from from small-scale laboratory tests, large scale ‘test rigs’, or full-scale plant operation?
- Is the available data good enough – is more extensive or better (higher resolution, better accuracy) data needed?
- Are the uncertainties in the measured data and test geometry understood and quantified?
- Does the data cover the region of parameter space of interest for the real application? Is the process of scaling the data to the real application well developed and robust, with the distortions it introduces well quantified?
- Has a similar model using the same or a similar code been validated against relevant data before? What evidence is necessary to provide sufficient reason to trust that the extent of the ‘similarity’ allows a new model to inherit the confidence from that validation evidence, without repeating it?
- Will a part of the data be used to calibrate the model, and so cannot be used for validation? If so, which parts of the data should be reserved for blind validation comparisons after calibration?

There are also questions related to the interpretation of comparisons to data:

- What is the appropriate validation metric, and what does obtaining a certain level of agreement with it imply about the confidence that can be placed in the model?
- For example, is the data representative of the most important model outputs, or is it a parameter that is convenient to measure, with some separation from the more important location? In this case, are there ways of showing how well that agreeing with this proxy measurement relates to agreement at the key location?

⁵ The ASME definitions for V&V are discussed Volume 1, Section 4.3, although others are in use.

Introduction and Technical Context

- Does the presence of intrusive instrumentation or test fixtures appreciably change the flow or phenomena, such that they should be physically modelled as part of the validation? Similarly, is the instrument operating in regions of strong gradients or fluctuations, and so does the dynamic or spatial averaging response of the measurement to these quantities need to be anticipated and replicated in the validation simulation?
- Is the data an integral quantity (a whole channel flow rate or mean temperature) that can show good agreement, while masking a disagreement with the details (e.g. the profile of velocity or temperature across a channel)? Similarly, is an extreme value being well predicted (such as a maximum component temperature), but the location where it occurs is either not known, or in disagreement? Being aware of the agreement at a level of detail able to test physical mechanisms⁶ reduces the chance of a result that agrees well by compensating errors ('being right for the wrong reason').
- Is the comparison able to be made at more than one operating condition? It may be possible to show good agreement with data for one combination of inputs (either by coincidence or via calibrated settings), where the effect of compensating errors or insensitivity to parameters at that condition coincide to cancel each other out, but the agreement diverges when moving to another operating point.

1.1.4 Uncertainty Quantification and Sensitivity Analysis

UQ and SA are related, but distinct, activities used to understand and quantify the implications of modelling uncertainties.

Uncertainty Quantification: Establishing the range of, and likelihood for, values that a simulation prediction could cover, which should (with an estimate of confidence) encompass the 'true' (but unknown) real value that the physical system that has been modelled would exhibit.

Sensitivity Analysis: Identifying how the uncertainty in a model output can be apportioned to different sources of uncertainty in its inputs⁷, and hence where efforts in terms of validation and UQ should be concentrated.

In this document, the term 'sensitivity assessment' will be used to describe activities in a broader sense and will cover less formal sensitivity studies, whereas SA will refer to the more formal and rigorous numerical analysis.

These topics will be discussed in significantly more detail in this volume, but high level questions that should be considered at the start of an analysis are:

- Are the key sources of uncertainty understood?
- What method should be chosen to assess uncertainty?
- Is the dominant uncertainty in model inputs, or in the model itself?

⁶ In NTH this is often called having 'CFD grade' data.

⁷ From the definition given by Saltelli *et al.* (2007)

Introduction and Technical Context

1.1.5 Overall Estimate of Confidence in a Prediction

All of the component activities outlined previously help form a case to answer the questions posed at the start of Section 1. Organisational approaches (examples of which are described in Section 3.1) provide mechanisms and frameworks assisting with, and progressing through, these decisions and activities. However, for any particular analysis task the choices made about each activity depend on the context and case specific requirements, the purpose of the model, and the onward use of the simulation results.

For example, if a conservative approach is adopted, then some SA and UQ activities may be required⁸. If such a conservative model is, for example, already a credible method endorsed by a design authority or regulator as suitable to demonstrate compliance with operational limits, then additional validation is unlikely to be necessary.

When a conservative approach is not possible, or is not appropriate for other reasons, then a suitable level of UQ is required. This will generically involve:

- Identifying the phenomena that need to be modelled.
- Identifying the specific output metrics that the model must predict.
- Identifying relevant sources of uncertainty, and their likely range.
- Identifying the sensitivity of the output metrics to these sources.
- Identifying how simplifications or incomplete descriptions of the model physics, and numerical errors in its solution contribute.
- Performing model evaluations to establish a probability distribution for each output metric.
- Assessing the implication of these probability distributions in the context of the problem being solved.

As noted in Figure 1.1, these do not need to be performed as sequential and linear steps. The 'output metric' is a specific property at a known location that is of key significance to the safety or performance of a plant or system. This can also be referred to as a 'quantity of interest', System Response Quantity (SRQ) or 'figure of merit'.

Of all the component activities identified, this volume aims to describe UQ in the most depth. Section 3 also provides context and sources of information on the other component activities.

1.2 Uncertainty

The conservative bounding approach to thermal hydraulic assessment is widely applied and can have the advantage that the models used are relatively cheap to develop and run. However, in some circumstances, they can result in excessive pessimisms that may lead to the prediction of an incorrect progression of events or unrealistic timescales of fault evolution. They can also predict behaviour so onerous that a safety case could not be made, requiring reassessment and reduction of conservatism. This can cause focus to be misplaced on identifying and addressing safety

⁸ Mainly what is necessary to demonstrate that an approach is actually conservative, which may not be apparent or able to be determined from the outset.

Introduction and Technical Context

issues that actually pose little risk. A corollary of this is that complexity in systems and models challenges the ability to know *a-priori* what conservative approximations are necessary, and have been shown to not bound the upper uncertainty band of BEPU predictions at all points of some transient simulations (Glaeser, 2008).

This has been acknowledged internationally, and since the late 1980s, there has been a move to making evaluations on a BEPU basis (Wilson, 2013). This approach uses a realistic model of thermal hydraulic phenomena and realistic input data to produce a best estimate output. Uncertainties in the inputs and modelling approach must then be accounted for when analysing the model results. This has the advantage that it provides a realistic view of how an operational scenario can progress, and an appropriate treatment of uncertainties ensures that an adequate margin is maintained between an uncertainty bound on the output metric and a safety limit. A realistic model has further benefits in that it provides better insight into real plant behaviour, which can inform its monitoring, diagnostics and maintenance.

However, UQ is not necessarily straightforward, and under-estimating the uncertainty could lead to a potentially unsafe or misleading result. The relationship between uncertainty, margin and BEPU vs conservative assessments is illustrated in Figure 1.2 (Figure 2.6 from Volume 1), and is discussed further in a licensing context in IAEA (2003).

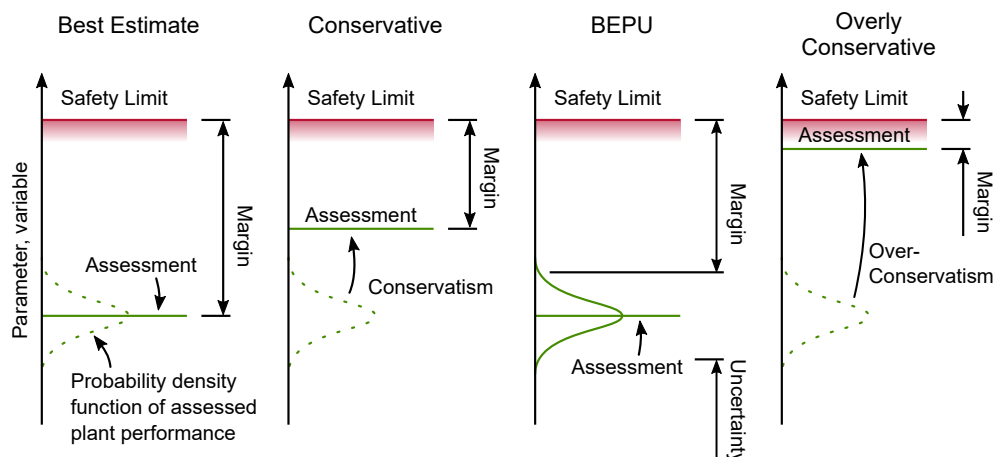


Figure 1.2: Margins and uncertainty in best estimate, conservative, BEPU and overly conservative assessments.

This technical volume provides an introduction to the methods applied and references to current international nuclear industry best practice guidance describing how to perform UQ in support of a BEPU approach to thermal hydraulic modelling.

Figure 1.3 illustrates an overview of the aims of UQ. For a given thermal hydraulic model different sources of uncertainty exist, including uncertainty in the model inputs and uncertainty in the model itself. When the identified sources of uncertainty are accounted for (for example by propagating a range of model inputs through the calculation route using range of numerical model options) the result is a distribution of the output metric of interest (in the example shown in Figure 1.3, a particular temperature).

Introduction and Technical Context

This can be used to determine the best estimate value⁹ of a temperature of interest, T_{BE} as well as a higher temperature where there is a calculated probability (with a known, numerical level of confidence) that this value, T_{Conf} , will be exceeded.

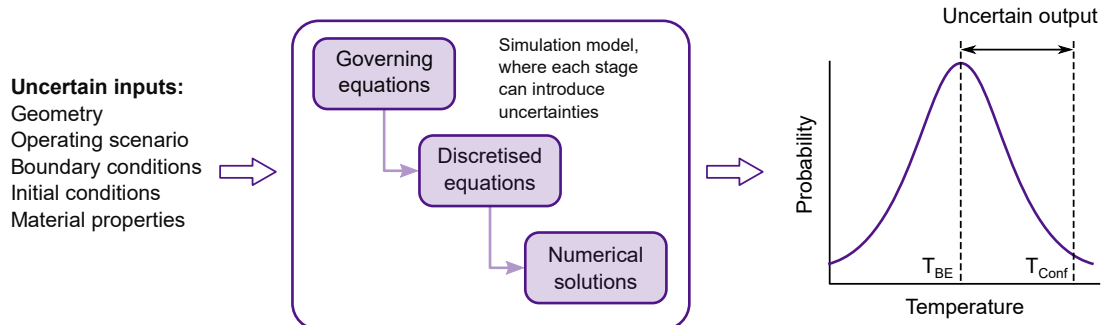


Figure 1.3: Uncertain inputs propagated through a model to produce a best estimate prediction and quantify the uncertainty in the prediction.

There are two types of uncertainty in any model:

Aleatory Uncertainty which occurs due to the naturally occurring random variability in physical processes and input parameters and therefore cannot easily be reduced. Aleatory uncertainty is also referred to as physical variability, irreducible uncertainty and Type A uncertainty.

Epistemic Uncertainty is caused by a lack of or incomplete knowledge about some part of the physical system being modelled, due to a simplified modelling approach that approximates certain physical phenomena (or couplings between phenomena), or by the use of representative data, rather than plant or circumstance specific data. Epistemic uncertainty can be reduced through gathering more (relevant) data, by increasing understanding of the physical system or increasing the fidelity of the modelling approach. This type of uncertainty is also referred to as the state of knowledge uncertainty, subjective uncertainty, reducible uncertainty or Type B uncertainty.

Roy and Oberkampf (2011) give more detailed definitions of these two types of uncertainty. Separating uncertainties into aleatory and epistemic¹⁰ provides clarity on which uncertainties could be reduced through, for example, improved validation against more measurement data or by increasing the complexity or specificity of a model. However, gathering the most pertinent data or making model improvements may be sufficiently difficult or costly that it cannot be accomplished, and so delineating between these two types of uncertainty can be ambiguous.

The first step in UQ is identifying the sources of uncertainty in the analysis, typically as the uncertainty in the input data or in how well the model represents reality. Then, depending on the nature of uncertainty and numerical model, different methods exist for assessing the uncertainty in the outputs of an analysis.

For example, an uncertain input x may be characterised by a Probability Density Function (PDF),

⁹ In general the best estimate value is the distribution median. In this sketch, the mean, median and mode (most-probable) value coincide, because it is represented as a normal (Gaussian) distribution. In general, the distribution may be skewed, and these values will not be the same.

¹⁰ To help remember which type is which: *alea* is the Latin for a game of dice – ‘*Alea iacta est*’ (‘The die is cast’); epistemology is the branch of philosophy concerned with the theory of knowledge.

Introduction and Technical Context

$\rho(x)$. This PDF¹¹ would then be sampled and the chosen values used as the inputs for a simulation using the model (propagated) to inform the uncertainty in the output parameter of interest.

NTH simulations can be computationally expensive to evaluate, and so it may prove necessary to use methods which take results from a high-fidelity model to develop an approximate, lower-fidelity, surrogate model (also known as a reduced order or meta-model) that is significantly faster to evaluate. Surrogate models are a key tool for performing SA and UQ, and the process of creating them introduces a degree of approximation, which (ideally) is itself quantifiable, and should be accounted for when interpreting its results. They are also used when evaluation speed is necessary, such as when the model is embedded as a component in a larger simulation, used in design optimisation, or applied to real-time simulation or control, where evaluating the full model would be intractable, too time consuming, or too expensive.

1.3 Thermal Hydraulic Phenomena that Contribute Uncertainty

The fundamental physics and equations describing the thermal hydraulic phenomena that can be present in a NPP are generally well understood. However, making predictions of them with high accuracy and certainty is challenging for reasons that broadly fall into two categories:

Complexity in behaviour, with numerous overlapping and interacting processes.

Variability or unknown values for properties and conditions, with some processes sensitive to small differences and details.

The uncertainty that arises in modelling due to complexity and variability have a number of origins relating to physical processes and phenomena, and to their mathematical representations¹². This includes the following aspects:

Spatial Variation and Evolution of Flows: Flows in real systems often occur in regions such as large vessels and plena, where the details of the complex three-dimensional flow field must be accounted for. However, even when flows are essentially one-dimensional (such as in pipes and annuli), they are rarely fully developed over the regions of importance – the effect of entrances, bends, valves, junctions, expansions and contractions propagate for significant distances and interact with downstream components. This can affect heat transfer, pressure losses and mixing substantially. Most high quality data and correlations for these processes in pipes and components assume fully developed flow (although specific and common situations like bend-bend interactions are documented, such as by Miller, 2009). By combining them in a flow network, further uncertainty is introduced in addition to the inherent uncertainty in the component correlations.

When building a model, some knowledge of the upstream details are needed to provide realistic inlet and boundary conditions. Even if the propagating effects or the interactions between components are known, including them requires that an upstream representation of a given part of the system must also be modelled, with increased time and resources.

¹¹ Note that the integral of a PDF is a Cumulative Distribution Function (CDF), which varies from 0 to 1, and gives $P(X \leq x)$, the probability that a random variable X will be less than or equal to x . It is also common to encounter the Complementary Cumulative Distribution Function (CCDF), which is $P(X > x) = 1 - P(X \leq x)$, and is sometimes called the 'survival function'.

¹² There are additional numerical solution and user aspects that also contribute uncertainty, as discussed in Section 2.1.

Introduction and Technical Context

Unsteadiness and Instability of Flows: As a corollary to the effects of spatial variation and development described above, real flows are rarely in a true steady-state. Plants and equipment either undergo transients, contain vibrations and acoustic sources, can have slow cyclic behaviour, bifurcations, or can have localised unsteady flow (such as in separated or mixing regions). This can also include Fluid-Structure Interaction (FSI), which can modify the flow.

Correlations developed in idealised laboratory conditions can be based on true steady-state conditions. However, applying them in an unsteady flow involves an approximation and simplification to assume that conditions are quasi-steady (the local flow process adapt quickly to the changing driving forces).

Systems may interact with the variable natural environment and weather, or exhibit long term behaviours that only emerge intermittently. Spontaneous instabilities can also be generated in otherwise unchanging conditions (situations involving buoyancy are especially prone to this). Modelling methods may not be able to be provided with sufficiently realistic and detailed information, be run for long enough to resolve such transients, or have sufficient fidelity to include mechanisms that drive subtle oscillatory processes.

Turbulence: A specific manifestation of unsteadiness is turbulence. As introduced in Volume 1, Section 4.5.3 and elaborated in Volume 3, Sections 2.2.3 and 2.2.4, turbulence epitomises complexity in fluid dynamics. It must usually be modelled with approximations and simplifications, inevitably causing the omission of details and generating uncertainty (although there are often ways of quantifying the inaccuracy).

Material Properties: For most solids and fluids to be modelled in an NPP, it is usually possible to obtain the relevant thermophysical properties with an acceptable level of accuracy, in their design or as-built condition. For more novel materials proposed in new designs, however, applicable data may be scarce, and contain significant uncertainty, especially at elevated temperatures. In addition, the properties of all materials could be subject to variation in composition and characteristics as a result of additives, or as they age, with contamination or exposure to radiation and chemical reactions.

Therefore, the 'current' operational properties of materials in a plant, representing the moment when a modelling prediction is required, may not be known with the same high confidence, without specific tasks to sample, measure and control them. Particular difficulty exists for surface properties, such as roughness and emissivity – these depend not only on the material, but also on how a component was fabricated, and then on the exposure, corrosion, deposition and ageing that it has experienced. These processes are often not well understood or easily modelled, but their influence on frictional losses and heat transfer can be significant.

Introduction and Technical Context

Geometry: It is usually possible to obtain the geometry for components that are to be modelled from engineering drawings and CAD models. However, these represent the designer's intentions, and the actual realisation of the as-built plant will contain a level of deviation from these due to tolerances and manufacturing variability. The components may then experience changes from their 'cold' dimensions when they are at their operating (or long-term aged) conditions due to thermal expansion, mechanical loading, or the effects of irradiation. Such deformations and distortions may be subtle and inconsequential, or may lead to significant changes in dimensions (such as in bowing of fuel assemblies or the shrinkage of graphite) and in the presence of flow paths (such as gaps appearing between structural elements, resulting in new routes for bypass flows).

It is also possible for the presence or omission of small geometrical features to introduce significant changes to results. For example, how rounded or sharp the corners of a flow entrance or contraction is, and whether small lips or chamfers are present, have been found to introduce significant changes in flows for these components. The judgements and choices of the analyst can introduce the resulting uncertainty because such details can be retained or 'defeatured' (removed) as part of the necessary process of developing a simplified computational geometry. Similarly, such ridges or steps can be present in a real plant that are not present in a CAD model, either through manufacturing deviations, protruding seals, wear, deposition, or misalignments caused by assembly or reassembly after maintenance.

Multiphase Flows: When multiple phases are present, the possible complexity of flows containing droplets, free surfaces, films, boiling and entrained bubbles is extensive. In pipe flows, this includes transitions between stratified, annular, bubbly, slug and churn flow regimes. The resulting mass momentum and energy transfers between phases and at walls have been extensively studied, and are often encapsulated in parameterised empirical 'closures' or correlations. Therefore, much of the uncertainty involved in predicting multiphase flows is related to assessing the appropriate choice (and parameter range applicability) for a flow being simulated by a given correlation, which itself may have been derived in idealised, isolated conditions, and have its own uncertainty.

Coupled Multiphysics Effects: The simulation of most physical phenomena is interrelated – for example the evaluation of structural integrity depends on pressures and heat transfer as input boundary conditions. However, some phenomena are more tightly coupled and have strong and fast feedback interactions. FSI, where large surface deflections influence the flow field is one example. Another is nuclear heating – NPP simulations usually include some level of energy source either from fission reactions or from decay heat. Determining this heating is an involved assessment, that may need to be made with imperfect knowledge, and so the uncertainty in the prediction or specification of the magnitude and distribution of these heat sources is inherited by NTH assessments relying on them. This applies when the information transfer is one-directional and uncoupled (applied as volume or boundary energy sources to a model), but is exacerbated where feedback and coupled effects are present.

Coupled effects occur, for example where the predicted nuclear heating is subject to feedback from temperatures predicted by an NTH model, as a result of neutronic effects (such as Doppler broadening) and geometrical changes (such as an expanding core increasing the spacing between fuel elements). The uncertainty in the predicted temperature field, and the uncertainty in the models for reactivity are closely coupled. There are more extreme examples of coupled effects,

Introduction and Technical Context

such as reactivity-flow instabilities in Boiling Water Reactors (BWRs). These are problematic for a modelling-led design and subsidisation process because, unless modelling methods are carefully designed to capture these combined effects, they may not predict the presence of some phenomena. Where such phenomena are predicted, they can require a large number of models to be combined (for example, in predicting CRUD induced power shift as demonstrated by Jones *et al.*, 2018), each of which contributes its own uncertainty.

The extent of complexity and variability that arises in NTH models means that a level of approximation, assumption and simplification will always be necessary. Furthermore, expert judgement and prioritisation of effort, regardless of the fidelity of the modelling approach, will also need to be considered. This volume aims to provide guidance on the concepts, tools and techniques that can be used to systematically understand the effects of doing so.

1.4 Why are V&V and UQ Hard for Industrial Users?

The availability and maturity of formal V&V, UQ and SA methods is not commensurate with the limited extent to which they are routinely applied in industrial NTH situations. This has been noted by practitioners of BEPU modelling (Ivanov *et al.*, 2019). It is, however, unlikely to be simply because '*Real V&V and real UQ are a lot of work*' (Weirs and Winokur, 2017), although this is also true.

One reason may be because SA and UQ methods in particular have formal mathematical foundations that may seem unapproachable for most practising thermal hydraulics engineers. CFD at an academic and implementation level also has such formalism¹³. This is, however, not visible to most CFD users because the tools of industrial practice abstract and encapsulate these in an implementation that makes their presence hidden. Some occasions will require users to explore and better understand the more formal details, but the well developed nature of the tools for NTH modelling provides little necessity or motivation to question and consider the application of them during routine use. The near-consensus around a framework of popular methods¹⁴ means that it is not necessary for a user to choose from one of several applicable, but incompatible approaches. A user's skills and knowledge are then generally transferable between software packages and industrial sectors, and can be added to gradually, where the application of new modelling options can be learned and tested incrementally. Applying UQ does not have these properties; each analysis process requires an investment of time and self-education for non-specialists to implement it within its own conceptual structure before any benefits can be realised, and it is harder to transfer between or compare approaches. They require some specialisation and a long term commitment to the approach within an organisation to apply them effectively and consistently.

The framework for NTH modelling is created by the physical world (items of plant such as vessels, pumps and pipes, which are made of real materials) and its coherent mathematical treatment is enabled by the laws of conservation of mass, momentum and energy, and by dimensionless groups. UQ, on the other hand, is an almost entirely mathematical construct, and so does not have this unifying underpinning that engineers and analysts rely upon (often without realising the extent to which they do). While the tools of probability and statistics are common to and coherent across UQ

¹³ Present in, for example, operator splitting for pressure-velocity coupling, considering a mesh as a 'Sobolev space', the characteristics of discretisation schemes, the properties of linear solver algorithms etc. This is similar in FEA.

¹⁴ The finite volume discretisation and RANS turbulence modelling being the 'work-horse' methods.

Introduction and Technical Context

analysis methods, these methods are initiated from different underlying principles and assumptions, and so this commonality is not necessarily unifying from a user's perspective. Therefore, the guidance provided in the 'comprehensive frameworks' for V&V and UQ (such as those referenced in Section 3.2) appear to end-users as a long way from being a complete and pragmatic guide to their application – in many cases they are a starting point and an organising principle only.

As an example, advanced CFD engineers¹⁵ usually require to know a significant quantity of three elements to fully exploit the discipline: fluid mechanics; mathematics/numerics and computing/programming. Typically two out of three of these elements are taught in undergraduate courses, and the final element needs to be learned either in post-graduate study, or via industrial experience. Requiring UQ to be conducted routinely then adds an additional element that should be recognised. Mechanical engineers, who comprise a sizeable proportion of NTH practitioners, are also typically not formally educated, experienced or confident enough in the fundamentals of statistics and probability theory to easily recourse to their use. It is likely to be the case that the full range of knowledge and skills will not be vested in each individual, and so additional engineers, scientists or mathematicians, with more specialised UQ knowledge, will be needed to support and supplement analysis. Similarly, to specify and incorporate validation data into an analysis structure requires individuals that are more specialised in experimental work. In this way, applying UQ to a CFD or NTH analysis can be considered as turning it into a multi-disciplinary activity.

With this starting point, and size of the topic in mind, this technical volume intends to provide a clarity on the available and required thought-processes, attitudes, concepts and options for approaching the topic. Where it does provide firm guidance and recommendations for practice, they are necessarily in a limited number of well understood cases. Doing '*real V&V and real UQ*' will be involved and case specific.

¹⁵ Meaning users who are skilled and independent, sceptical and questioning, able to fully exploit and customise the code, understand the meaning of numerical settings and can choose (and know the limits of) models, discretisation settings etc.

2 Concepts for Uncertainty

This section describes the concepts and introduces the mathematical tools used to quantify uncertainty and perform sensitivity analysis. UQ and SA, while not the same, are closely related – they share concepts and methods and can be performed simultaneously in some cases. UQ addresses the statistical description of input uncertainties and their propagation through a model, focussing on quantifying the impact of this uncertainty on model outputs. SA on the other hand considers which elements of the uncertainty contribute the most to variance in model outputs. Typically, a simpler set of sensitivity analysis will be performed early in the NTH model development life-cycle, and used to inform UQ.

The equivalent conceptual and mathematical underpinnings for V&V are not covered; the required methods for V&V are generally well documented or referenced alongside specific instances of guidance and standards.

2.1 Sources and Treatment of Uncertainty

When performing UQ or SA on a model of a physical system, the first step is to identify possible sources of uncertainty. The process of determining what uncertainties exist may be a significant undertaking, and can indicate many more sources than can be practically assessed.

2.1.1 Common Sources of Uncertainty

Figure 2.1 shows several types of sources of uncertainty that will commonly need to be considered in an NTH model.

The uncertainty in the geometry, materials properties and status or configuration of the plant or equipment being modelled can be grouped together as uncertain **model parameters** (often referred to as parametric uncertainty). Related to this is uncertainty in the **initial and boundary conditions** for the plant, or sub-systems – these could arise because of natural variability in, or lack of knowledge about, the flow rates, temperatures, pressures and turbulence at the (user selected) boundaries of the system being modelled, internal heat generation sources, or heat transfer to external or ambient conditions.

Model uncertainty (also referred to as structural uncertainty) arises because the mathematical model implemented in the code is only an approximation of the physics in reality, and often contains deliberate idealisations.

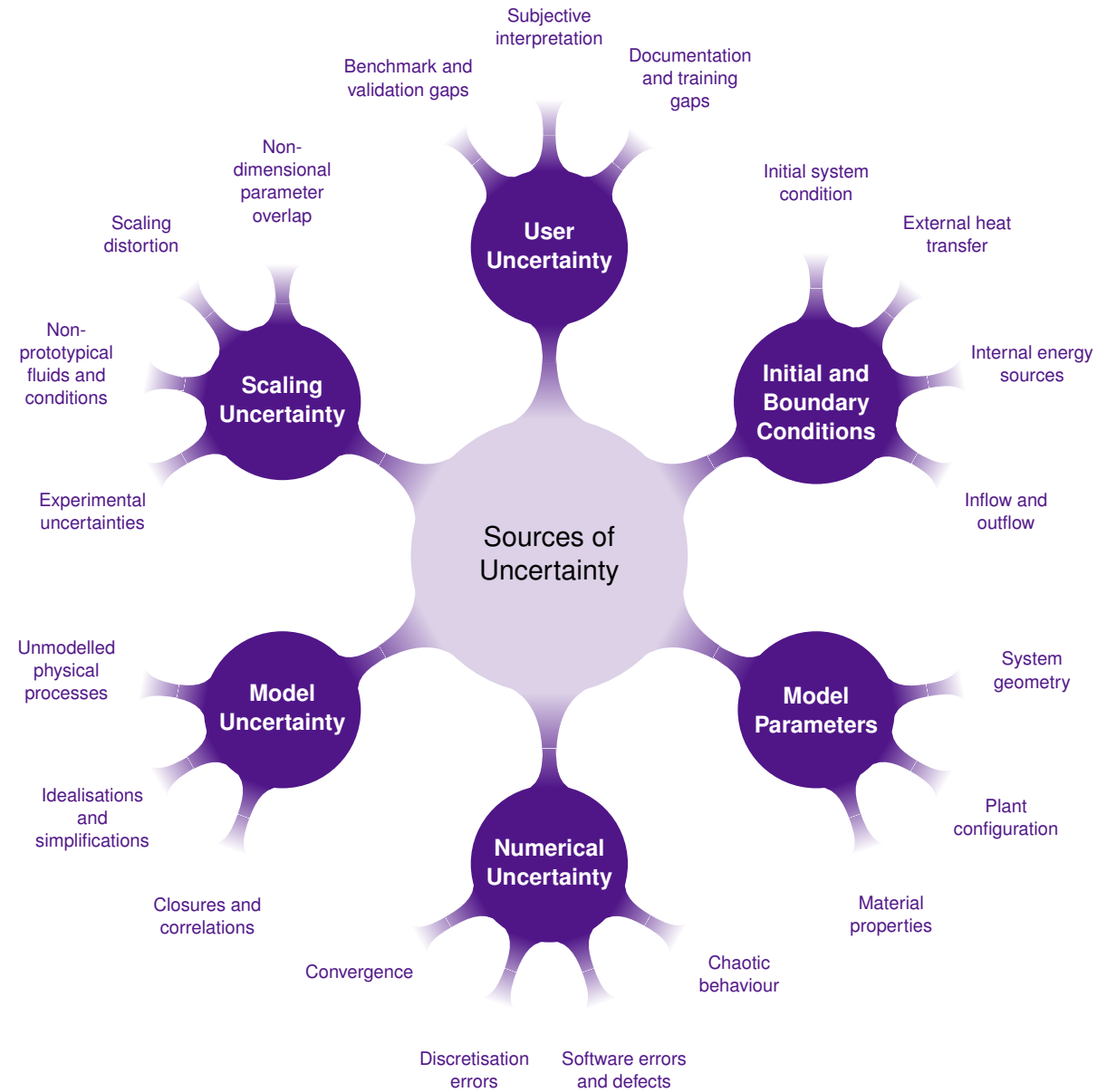


Figure 2.1: Sources of uncertainty in a thermal hydraulic model.

Concepts for Uncertainty

Thermal hydraulic models consist of balance equations for the conservation of energy, mass and momentum, as well as models for material properties and closure and constitutive laws. System codes rely heavily on empirical correlations (closures) which are models that represent complex phenomena by a simplified relationship between transport fluxes of energy, mass or momentum and local flow conditions. The application of Reynolds-Averaged Navier-Stokes (RANS) turbulence models (which are also closures) in CFD produces a similar source of model uncertainty.

In addition to being a source of minor inaccuracies, model uncertainty could lead to a model being substantially erroneous or inadequate for the desired task. A model may deliberately omit, or be incapable of representing a significant phenomenon. This could arise because the presence or importance of that phenomenon was not expected and anticipated when the modelling process began, or the model may not contain a sufficiently detailed or realistic representation of the physical processes necessary to predict it.

For all models, the user is required to decide which phenomena need to be modelled, determine an appropriate representation or simplification of a real geometry, including the overall size of the domain, how to include individual parts and what initial boundary conditions to apply. Figure 2.2 highlights the importance of **user uncertainty** in the results of NTH calculations. In both cases presented, different organisations were asked to perform blind predictions of a benchmark experiment. The results of these exercises showed significant spread in the predicted metrics between the different participants, much of this spread may be associated with user uncertainty. Differences in model set-up, including the nodalisation (choosing types of component and their spatial discretisation) as well as interpretations of boundary conditions and modelling choices contributed to the user uncertainty. User training, best practice guides and this kind of benchmark and validation against experimental data can help to reduce this source of uncertainty in model outputs. User uncertainty is also linked to the occurrence of unmodelled physical processes, because the identification and ranking process in creating, say, a Phenomena Identification and Ranking Table (PIRT) (Section 3.1.1) is inevitably subjective, and the extent of the uncertainty caused cannot be assessed.

Scaling uncertainty arises from the fact that model or sub-model validation and correlation development is often performed on reduced scale experiments using substitute materials or obtained from operating experience at a different scale. Several aspects of the real system behaviour are scaled to make experiments feasible, cheaper to conduct and easier to instrument. For example, by using reduced power, size, pressure or temperature, or by using simulant fluids (CSNI, 2017b). The purpose of a scaled experiment is to match the most important dimensionless quantities that characterise the real system. However, NPPs contain such a range of phenomena and interactions that it is not possible to simultaneously preserve all length scales and time scales, along with the flow and heat transfer regimes. This is known as ‘scaling distortion’. It is also possible that the real system being modelled operates in regions of the dimensionless number envelope not covered by the experiment, or that the experimental geometry is not exactly representative, and so the adequacy of the validation evidence for supporting models of the real system must be assessed. Thus, there may be a requirement to perform extrapolation outside of the range of experimental parameter space, which leads to uncertainty. Furthermore, experimental results contain uncertainties themselves, in the accuracy or location of their measurements and processed data, but these are not always well documented and characterised.

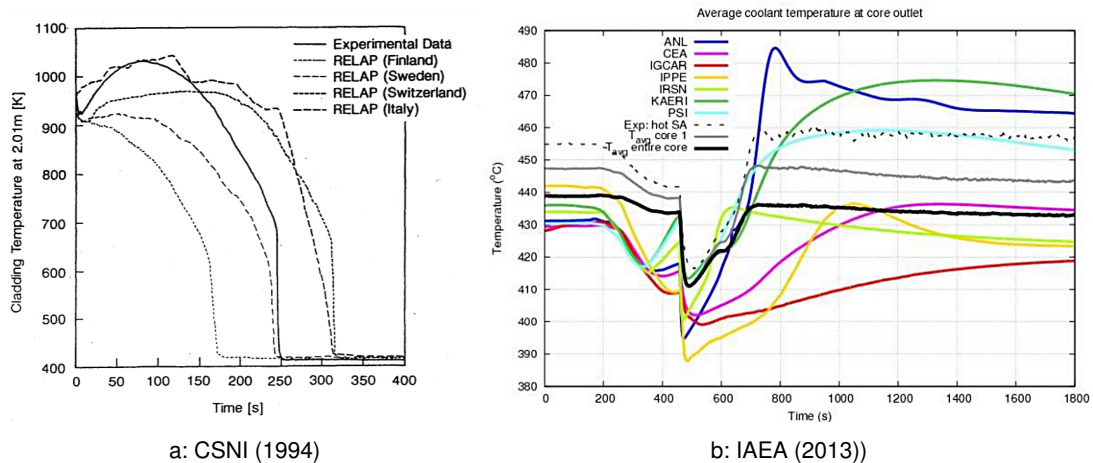


Figure 2.2: Examples of user effects in system code calculations.

a: Comparison of four RELAP5 model calculations of cladding temperature for CSNI International Standard Problem number 25 (end of accumulator discharge in a postulated LOCA) to experimental data. The wide range in predicted cladding temperature and quench times from the different modelling groups are indicative of considerable user uncertainty.

b: Comparison of blind modelling predictions for the Phénix prototype pool-type sodium-cooled fast reactor end-of-life tests programme. The spread in predictions of coolant temperature at core outlet are indicative of the uncertainty in modelling predictions.

Numerical uncertainty arises through all of the processes involved in implementing and solving a numerical model. In CFD models, poor mesh quality or inadequate geometrical or time resolution can contribute. This is in addition to the truncation errors and temporal, spatial and equation¹ discretisation errors intrinsic to the computational method, as well as incomplete or unstable equation solver convergence. All of these lead to potential uncertainties in the model calculation results (that are hard to quantify in realistic calculations).

Some physical systems exhibit complex unstable chaotic or periodic behaviour, and numerical methods may or may not be able to correctly capture this. Likewise, some numerical methods exhibit unstable solutions which are a result of purely numerical features – telling the difference and having confidence that the predicted presence or absence of complex behaviour is ‘real’ is possible in some cases, but requires specific effort.

Numerical software tools² are elaborate, and are generally not completely free of implementation errors and defects (‘bugs’) – these can manifest themselves in ways that (subtly or grossly) distort the results of a calculation.

¹ Meant in the broad sense, of converting Partial Differential Equations (PDEs) into algebraic linear systems for solution, using the finite difference, element or volume method, as well as numerical schemes for evaluating gradients and interpolation.

² As well as the compilers and libraries for the code that they are written in, the operating systems that they run on, and the hardware that executes them.

Concepts for Uncertainty

2.1.2 Treatment of Different Types of Uncertainty

The way that uncertainties are selected and characterised depends on how UQ will be incorporated and treated in the overall prediction and decision making framework, and on the nature of the output metric being considered.

Uncertain inputs to be propagated for UQ can generally be treated by one of three methods:

Continuous probabilistic treatment: The most common, and most straightforward approach for propagating uncertain inputs through a model. Here the uncertain input, x , is represented by a PDF and its associated parameters. This describes the likelihood of all different possible values of the uncertain input occurring. A continuous probabilistic treatment is typically used to describe aleatory uncertain inputs (or epistemic uncertain inputs which can be treated in an probabilistic manner).

Interval treatment: Where an uncertain input is bounded by a known maximum and minimum interval range, the uncertainty is defined by the ends of an interval $[a, b]$ which contains the uncertain variable x . When propagated through a function or model the uncertain output metric is contained within a new interval $[c, d]$. An interval treatment is commonly used to represent epistemic uncertainties.

Discrete treatment: Uncertain inputs can also be represented by a discrete set of options, each with a specific probability of occurrence and such that the sum of all the options is 1. For example, a CFD analysis may be sensitive to the choice of turbulence model, and the modelling engineer could choose a most likely turbulence model, but other possible models should be considered in the UQ, each with their own probability of occurrence.

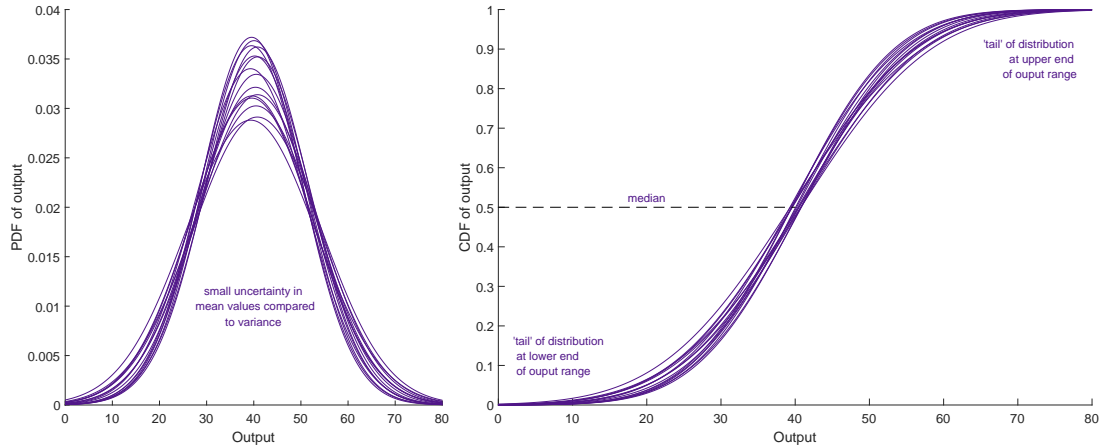
It is normally desirable to represent as many uncertain inputs as possible via continuous probability distributions. Incorporating discrete probability options or intervals into the UQ can lead to large increases in the number of possible model calculations performed, for example, requiring multiple separate sets of samples for each combination of epistemic uncertainties.

2.1.3 Significance of Output Metrics

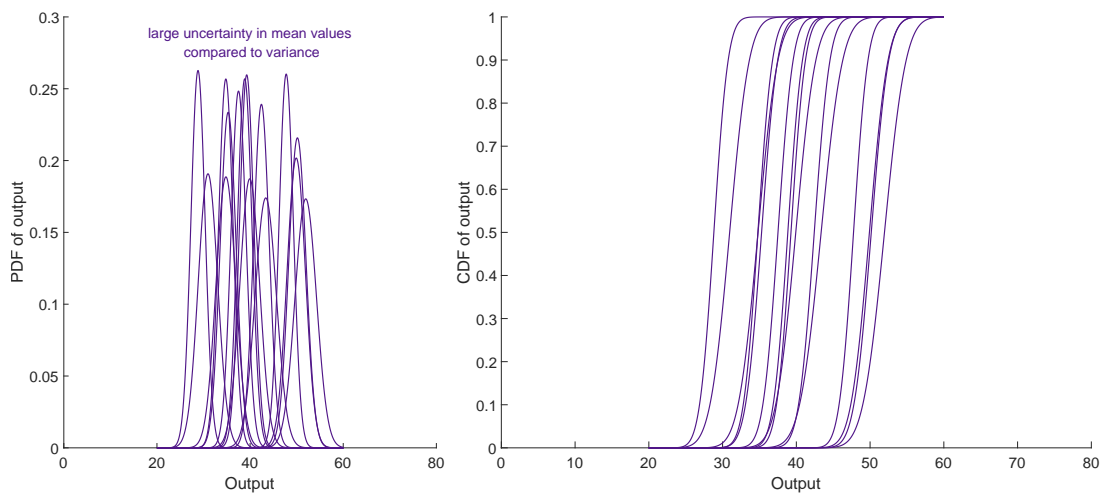
The way different input uncertainties are treated also depends on the nature of the output metric of interest. Some modelling predictions require the 'mean' behaviour of the system to be characterised. This requires the epistemic uncertainty to be well defined to give confidence in the output metrics. Other types of analysis may be concerned about the 'worst-case' behaviour of the system, which may be dominated by the low probability 'tails' of uncertain inputs which contribute to the aleatory uncertainty. For example:

- Turbulent or unstable flows produce mean loading, or heat transfer on a structure, but the details of the spatial and temporal spectrum of the fluctuating or transient part may induce a non-linear response in damage or fatigue loads, caused by infrequent combinations of fluctuations.
- For a large number of canisters storing hazardous materials, only a small number of them may reasonably be expected to fail, but these constitute the main component of the hazard. Those which do fail will have been subject to the most onerous loading or conditions, whilst also having the least integrity as a result of lower structural material thickness, or from more

weld defects. The mean behaviour of the canisters may represent no concern, but predicting a failure probability or rate of the outliers accurately will depend on knowledge of the combinations of low probability extremes for several uncertain inputs³.



a: Large variance, but smaller uncertainty in the mean, and so aleatory uncertainty dominates.



b: Large epistemic uncertainty on the mean, and so this dominates.

Figure 2.3: Examples of PDFs and their integrals (CDFs). In each situation a, or b, 15 normal distributions with a range of means and variances (widths) are plotted. Adapted from Vose (2008).

Figure 2.3 shows how the uncertainty in an output metric differs in the scenarios where:

- The epistemic uncertainty is small and the aleatory uncertainty is large. If the worst-case behaviour of the system is of interest and aleatory variability is large, then this may dominate the output metric uncertainty. As a result, detailed investigation of small epistemic uncertainties may not be required.
- The epistemic uncertainty is large and the aleatory uncertainty is small. If the mean behaviour of the system is the key output metric, then it may be possible that a full treatment

³ A specific example of this is discussed by Jiang *et al.* (2021) for TRISO fuel particles, where failure of only a small number within a pebble represents the dominant potential radiological source term, and modern modelling methods are able to be applied to assess the statistical likelihood of such failures

of the aleatory uncertainty is not necessary. Instead, effort should be focused on epistemic uncertainties which impact the average value of the output metric.

These cases represent clearly distinguishable situations, but often in realistic cases, the encountered behaviour will be somewhere in between, and considering a mixture of epistemic and aleatory uncertainty will be required.

2.1.4 Initial Sensitivity Assessment

For both model parameter uncertainty and uncertainty in the initial and boundary conditions, initial sensitivity assessments can be performed to highlight the importance of an uncertain input to the problem being modelled. If the simulation result proves to be not sufficiently sensitive to the expected (or estimated) range of changes in uncertain inputs, then resources likely do not need to be spent propagating their uncertainties through the model. Reducing the number of uncertainties to propagate is desirable to reduce the effort expended in the uncertainty analysis. Performing these assessments is also helpful in identifying important model characteristics such as the smoothness of the response, non-linear trends and discovering how robust the solution process is. As a result, often an initial sensitivity assessment will be performed to make a preliminary evaluation of relative importance prior to embarking on an extensive set of simulations or detailed UQ.

Although initial sensitivity assessment is exploratory in nature, the following high level activities should be considered during the process:

- Identify a list of key sensitivity analyses to perform. This is likely to include a degree of engineering judgement and the use of organisational approaches such as PIRT (Section 3.1).
- Define the output metrics to assess the parameter sensitivities against. These are likely to be the same output metrics for the full NTH model, but this is not a necessary requirement.
- Identify a physically reasonable upper and lower bound of each uncertain input parameter to perform the initial sensitivity assessment for.
- Run model calculations for combinations of uncertain inputs. As a minimum, sensitivities should be run at the upper bound, lower bound and central values of the input parameter. Generally, this will be for the full thermal hydraulic model, but it may be possible to run a simplified version of the model or separate calculations to reduce the computational effort.
- Quantitatively rank the input sensitivities for each output metrics of the model.

Based on the rankings a judgement can be made on whether or not to include an uncertain input in further, more detailed, calculations. These initial sensitivity assessments can thus help determine the key uncertain inputs early in the modelling life cycle and optimise how resources are directed. For uncertainties not carried forwards the reasons behind this should be documented and justified. The methods introduced in Section 4.5 are suitable, although applying only subset of them in an approximate/informal way will be appropriate, given the typically rapid and preliminary nature of an initial assessment.

2.2 Common Methods and Tools

2.2.1 Basic Statistics

To make use of the concepts, methods and tools described in this technical volume, it is expected to be highly beneficial for practising NTH engineers to increase their knowledge of (or refamiliarise themselves with) the fundamentals of statistics, or to have access to the advice of suitably specialised individuals. This is necessary to avoid misleading oneself or others in the interpretation of the outputs from the suggested methods, to create appropriate model inputs, and allow rigorous and robust arguments to be made about the statistical significance of results.

Error propagation: It is expected that some inputs to NTH models will come from experimental results, or will be characterised by non-dimensional parameters containing geometry, flow and material property characteristics. This is also the case for interpreting experimental data for validation comparisons. When applying UQ to both of these tasks, an assessment of combined errors or uncertainty in derived quantities is needed. The propagation of errors through simple algebraic expressions is introduced in Section 4.2.4 and well documented – for example NIST (2021) covers the fundamentals and Ku (1966) and JCGM (2008)⁴ provide more detail.

Probability distributions: It is common, in the absence of other knowledge, to assume⁵ that data can be represented by a normal (Gaussian) distribution. There are large number of distributions available with a range of applications and origins, and NIST (2021) describes common ones. Each distribution is characterised by ‘moments’ the first four of which are mean, variance⁶, skewness (extent of asymmetry) and kurtosis (how ‘heavy’ or ‘light’ the tails are).

Statistical hypothesis testing: Where a confidence interval (Section 4.7) can be assessed for UQ or validation purposes, then a statistical test can be applied. This allows the likelihood that an observed outcome could be the result of chance to be quantified, and how the certainty of statements relates to how much data is needed to assess them. NIST (2021) describes these, and the application of the concepts will be described in introductory statistics textbooks (e.g. Navidi, 2020).

2.2.2 Toolkits

In developing this technical volume, it was concluded that there is insufficient literature available that is basic, introductory and explanatory enough to make UQ methods accessible from first principles to the majority of practising or prospective NTH engineers and analysts. To start from the fundamentals would require a relatively extensive period of study, or access to a specialist already experienced in this domain. Many of the most succinct, accessible and clear available sources of information that introduce V&V and UQ are presentations where the topic is explained by example (such as Weirs and Winokur, 2017 or the Dakota Software Training, 2016), not formal papers or books. This is a similar position to using CFD, where having access to an experienced practitioner to ask for their guidance is the best way to learn, and in-lieu of that, or to extend into a new capability area, much of the best guidance on practice is in the form of tutorials and training

⁴ Available at www.bipm.org/en/publications/guides/#gum, and also referred to as ISO/IEC GUIDE 98-3:2008.

⁵ sometimes unconsciously, or without realising that this assumptions is implicit in a concept or method.

⁶ standard deviation is the positive square root of the variance

materials.

Therefore, a main recommendation is that the establishment or extension/expansion of a capability in UQ should be based on the methods already implemented in available tools⁷. This is greatly facilitated by several tools that are used and relied upon in practice in the nuclear industry being open source, such as Dakota⁸, OpenTURNS⁹ or Uranie¹⁰. Commercial UQ (and more broadly, optimisation and design exploration) tools are available, and more general-purpose numerical tools such as MATLAB¹¹, Python¹² or R¹³ also contain many of the necessary building-blocks with (often open source) libraries providing specialised surrogate model or UQ capabilities (Bouhlef *et al.*, 2019, Tennøe *et al.*, 2018). Their associated documentation, training materials, and user communities are likely to be the best method for establishing a UQ capability or adding to its robustness, acceptance and transferability/comparability. However, even with this approach, a certain level of understanding of statistics will be necessary to interpret the outputs meaningfully.

These tools typically run as an overarching code, and incorporate Design of Experiments (DoE) tools (Section 4.3). They request/initiate the specified evaluations of a separate analysis tool, then build surrogate models and/or calculate statistics of the collated outputs. This allows a UQ and SA toolbox to be overlaid onto existing analysis tools and models, that users will already be familiar with. An additional advantage is that the software that is available for UQ and SA is also usually capable of optimisation and model calibration¹⁴, providing a powerful set of methods for design and interpreting or exploiting experimental data.

2.2.3 Visualisation

Choosing an appropriate visualisation method can significantly aid the interpretation of whether a sufficient number of simulations have been performed for a UQ or SA analysis, the adequacy of their coverage of the input parameter space, and what the response of a model has been. This is especially true in the earlier, exploratory stages of an assessment. It is possible to use purely numerical or statistical metrics to interpret UQ or SA, if the behaviour and response is well known, but without visualising the data, important features will not be apparent. For example, if an output is bimodal (has two peaks or clusters of results, for example caused by a bifurcation of a physical process) then the mean value will lie somewhere between them, and represent neither well – simply assessing the mean and standard deviation would miss this insight.

The distribution of outputs matters, therefore, a range of visualisation methods should be employed:

Scatter plots: The relationship between two variables (inputs or outputs) can be observed as a plot of samples. The extent or absence of correlation between them can be seen in the shape of the cloud of points, and they can be used to interpret reason for the calculated value of a correlation coefficient (Section 4.5). Scatter plots are often shown as a matrix of all inputs

⁷ This is perhaps inevitable, particularly if made analogous to an organisation establishing a CFD capability – it would not be reasonable or sensible to expect to do anything other than choose an existing code and 'learn by doing'.

⁸ dakota.sandia.gov

⁹ openturns.github.io/www

¹⁰ sourceforge.net/projects/uranie

¹¹ www.mathworks.com

¹² Via NumPy numpy.org and SciPy www.scipy.org

¹³ www.r-project.org/

¹⁴ Table 3.1 defines this and distinguishes between it and other related activities.

against all outputs (or every input and output against each other, giving a symmetric square matrix) to allow the whole dataset to be surveyed.

Bar charts: Correlation and sensitivity coefficients as well as Sobol indices are best interpreted as how the size of one relates to that for other variables – bar charts are well suited to this.

Histograms: The distribution of data is indicated by the number of samples found in discrete ‘bins’.

A continuous version of this is a kernel density plot (Section 4.2.4).

PDFs and CDFs: Fitting the histogram of data to a known form of probability distribution is a useful way to characterise, compare and plot results. The integral of a PDF is the CDF, which provides a direct visualisation of what proportion of the data is lower than a given value – this allows confidence intervals or likelihoods of occurrence to be observed directly. A CDF can also be created numerically as the sum of a histogram or integral of a kernel density function.

Box and violin plots: A box plot shows the median and interquartile range (from 25 % to 75 %) of a distribution of data, as well as its minimum and maximum range. By combining features of a box plot and a kernel density plot, violin plots show all of the details of a distribution, as well as the quartiles of the data, and are particularly useful when comparing several datasets.

Creating several of these plots from the same dataset is convenient with common numerical tools. They are also the visualisation methods that are broadly applied in statistics and data science, and open source tools are available, for example Plotly¹⁵ or seaborn¹⁶ in Python.

Examples of PDFs are shown in Section 4, the use of scatter plots, histograms, bar charts and CDFs is demonstrated in Study B, and an example of a violin plot is shown in Figure 2.4¹⁷.

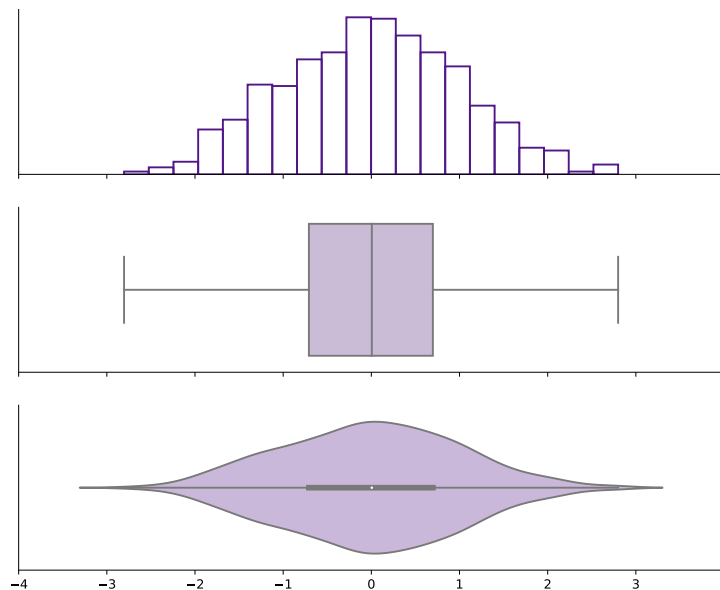


Figure 2.4: 1000 random samples drawn from a standard normal distribution visualised as a histogram, box plot and violin plot.

¹⁵ plotly.com/graphing-libraries

¹⁶ seaborn.pydata.org, which is based on Matplotlib (matplotlib.org) and described by Waskom (2021)

¹⁷ Adapted from towardsdatascience.com/violin-plot-its-time-to-ditch-the-box-plots-785629b0ff3a

2.3 Propagation of Uncertainty

Once the uncertain inputs of interest have been selected, the next step is to propagate¹⁸ them through the model to understand their impact. The overall aim (summarised in Figure 2.5) is to quantify the uncertainty in the model output metric of interest, often in the form of a PDF.

At the outset of an assessment, it is necessary to choose a conceptual framework for the application of statistical methods. There are broadly two methods in statistical analysis – frequentist and Bayesian. Section 4.1 provides a high level overview of these two approaches with respect to the subject of UQ to provide context for the topics covered in the remainder of the section.

To propagate uncertain inputs through an NTH model, they need to first be appropriately characterised. This involves understanding, for example, where the uncertainty originates from, whether it is aleatory or epistemic and how the uncertainty can be incorporated into a mathematical framework. A mathematical description of each uncertain input can then be defined. This is discussed in more detail in Section 4.2.

To assess the impact of a set of uncertain inputs on an output metric, an NTH model needs to be evaluated at a number of points in the parameter space to capture the underlying model response to these uncertainties. The range of parameter space and number of evaluations required is determined by the characterisation of the uncertain inputs and the type of uncertainties considered. Defining the points in the parameter space at which to evaluate the model is known as sampling – a range of methods for which are discussed in Section 4.3.

In practice, high-fidelity NTH models are often computationally intensive and executing them multiple times to propagate uncertainty can be

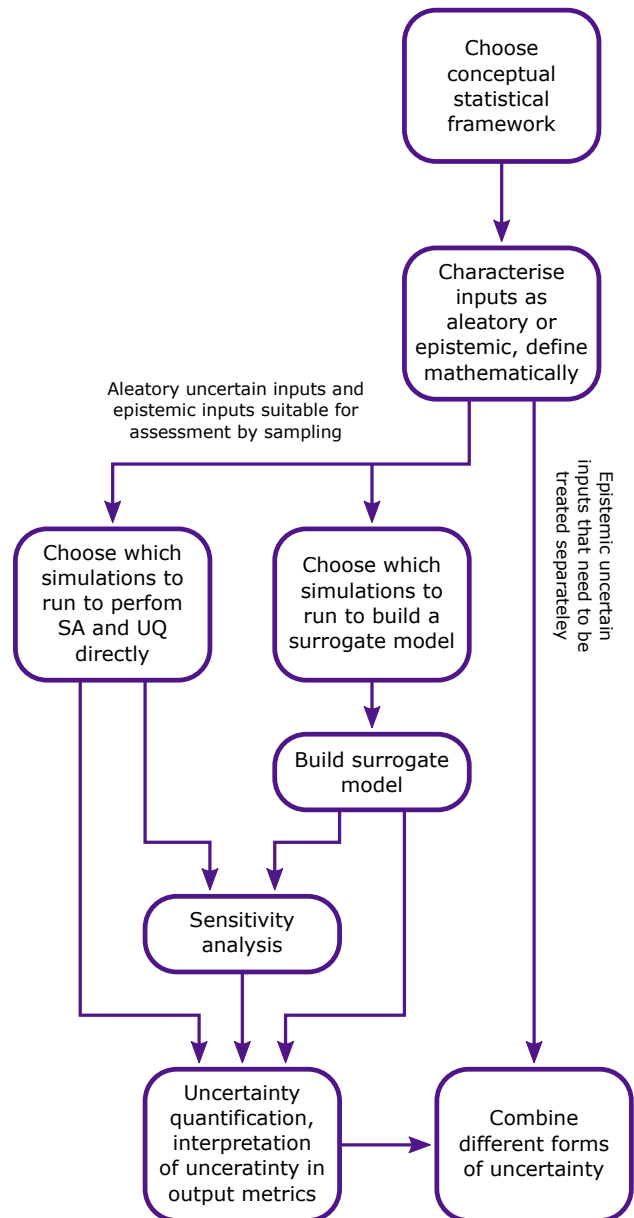


Figure 2.5: Assessment process and topics detailed in Section 4.

¹⁸ The alternative to propagation methods are 'extrapolation' methods (CSNI, 2016b, Section 5), which are generally applied to the scaling of IET data. Most UQ methods use propagation, and they will be the focus of this volume. It is also possible to apply Inverse Uncertainty Quantification (IUQ) where knowledge of the output is used to determine uncertainty on the inputs. This technical volume only considers forward propagation, although IUQ is starting to be applied to system code assessments (Section 3.3).

prohibitively resource intensive. Instead, a surrogate model can be developed, based on the model outputs of the high-fidelity model, to allow repeated interrogations of the surrogate model to propagate the sources of uncertainty. Care needs to be taken when developing these surrogate models, because a poor representation of the full-order model could distort or contribute significantly to the uncertainty in the model outputs. Section 4.4 discusses the requirements of a surrogate model and provides an overview of different types that can be applied.

Sensitivity analysis draws upon the techniques outlined above to determine which model inputs have an important effect on the model outputs of interest. The focus is on identifying the most important phenomena, model parameters and initial and boundary conditions, and which variables and relationships capture this. It can be performed at various stages in development of a model. Section 4.5 describes the process for performing SA and outlines useful metrics that an NTH analyst can use to assess the parameters to which an output is most sensitive.

Section 4.6 draws on the techniques mentioned above to describe how UQ can be performed, Section 4.7 discusses how its outputs can be interpreted, and Section 4.8 describes how to combine this with other forms of uncertainty.

After reading Section 3, it is recommended that Section 4 should be read in its entirety once, with the expectation that it will contain a large range of potentially unfamiliar concepts, whose relevance and applicability is not immediately obvious. After subsequently beginning to apply UQ tools and methods to their own problems, the intention is that readers will recognise the terminology and options for methods, and can return to the relevant part of Section 4 for guidance and to obtain the suggested references.

3 Methodologies

3.1 Organisational Approaches

Organisational approaches are primarily intended to transparently record and present a common view that has been elicited from a wide range of parties, highlighting gaps and facilitating communication and tracking of progress and status. The most common approaches are introduced in Sections 3.1.1 to 3.1.3. They are typically originated by national research bodies, regulators and international nuclear industry agencies (such as the OECD NEA and the IAEA) who commission and publish international studies of their use, some examples of which are given in Section 3.1.4, describing where they have been applied in practice.

3.1.1 EMDAP, CSAU and PIRT

One of the most established approaches is the Evaluation Methodology Development and Application Process (EMDAP), which is described in the United States Nuclear Regulatory Commission's (US NRC) Regulatory Guide 1.203 (US NRC, 2005). It is intended to provide a process, suitable for licensing, to develop and assess 'Evaluation Models' (numerical tools) used to evaluate accidents and transients. It contains 20 steps, grouped into four elements:

1. Establish Requirements for Evaluation Model Capability.
2. Develop Assessment Base.
3. Develop Evaluation Model.
4. Assess Evaluation Model Adequacy.

One of its required steps in Element 1 is creating a PIRT, and a UQ process is required in Element 4. The expected (or default) UQ methodology, typically applied to system code analysis, is Code Scaling, Applicability and Uncertainty (CSAU) (US NRC, 1989, Wilson, 2013). The Office for Nuclear Regulation (ONR) also consider CSAU as relevant good practice, and a suitable reference method to compare to (ONR, 2019).

An introduction to PIRT and the steps involved is given in Volume 1, Section 4.2, and while the process was developed as part of CSAU, it has grown to become a powerful tool in its own right (US NRC, 2005):

The PIRT should be used to determine the requirements for physical model development, scalability, validation and sensitivities studies. Ultimately, the PIRT is used to guide any uncertainty analysis or in the assessment of overall [evaluation model] adequacy. The PIRT is not an end in itself; rather it is a tool to provide guidance for the subsequent steps.

It should be kept in mind, that it contains subjective opinions, and is only a part of an overall, iterative evaluation process. For example, a PIRT can be used to develop Separate Effect Tests (SETs) and Integral Effect Test (IET) facilities, which themselves form part of adding to the necessary knowledge for an evolving PIRT for an NPP.

It is also possible for activities like scaling and sensitivity analysis to help to produce quantified ranking in a PIRT (Martin, 2011, Yurko and Buongiorno, 2012), rather than using only qualitative assessments, such as assigning a low, medium or high importance to an identified phenomenon.

3.1.2 Predictive Capability Maturity Model

Another common method is the Predictive Capability Maturity Model (PCMM) (Oberkampff *et al.*, 2007), originating from Sandia National Laboratory. It can be applied to help manage risk and avoid errors in any Modelling and Simulation (M&S) activity that is used to inform a decision, and asks a group of stakeholders to score the maturity of six elements:

Representation and Geometric Fidelity	What features are neglected because of simplifications or stylisations?
Physics and Material Model Fidelity	How fundamental are the physics and material models and what is the level of model calibration?
Code Verification	Are algorithm deficiencies, software errors, and poor Software Quality Engineering (SQE) practices corrupting the simulation results?
Solution Verification	Are numerical solution errors and human procedural errors corrupting the simulation results?
Model Validation	How carefully is the accuracy of the simulation and experimental results assessed at various tiers in a validation hierarchy?
Uncertainty Quantification and Sensitivity Analysis	How thoroughly are uncertainties and sensitivities characterised and propagated?

A maturity score of 0 to 3 for an element is based on the suitability of that aspect of a modelling capability to play a role with an increasing given level of impact or importance:

0. Low Consequence, Minimal M&S Impact, e.g. Scoping Studies.
1. Moderate Consequence, Some M&S Impact, e.g. Design Support.
2. High-Consequence, High M&S Impact, e.g. Qualification Support.
3. High-Consequence, Decision-Making Based on M&S, e.g. Qualification or Certification.

The elements are usually presented as the rows of a table format and the maturity levels as the columns. Weirs and Winokur (2017) describe the intent of PCMM as being to:

- Clarify the expectations of the customer, stakeholders and project team with respect to the rigour needed in a computational simulation study.
- Ensure that a consistent set of technical questions are asked of the team.
- Illustrate progress over time, across broad technical fronts.

Experience of the use of PCMM in practice brought about an evaluation of the utility and coverage/emphasis of its original formulation, and it has evolved to a ‘fourth generation’ form (Hills *et al.*, 2013). As part of its application to the Consortium for Advanced Simulation of Light Water Reactors (CASL) programme, comparisons were made between PCMM and CSAU, and it was concluded that neither is necessarily considered complete, but are not fundamentally different or incompatible, and applying a method that combines aspects from both may be a beneficial approach (Rider *et al.*, 2013). In a similar way to PIRT, methods are also under development to apply quantitative assessments of the maturity (Lin and Dinh, 2020).

3.1.3 Approaches from Other Industries

There are also organisational approaches that have been developed in other industries to address similar challenges of the increase in importance of modelling and simulation. For example:

- Quantification of Margins and Uncertainties (QMU) is used in the US nuclear weapons programme (Pilch *et al.*, 2006, Helton, 2009), and has commonalities with PCMM in key authors from Sandia National Laboratory.
- Aerospace is another regulated, high consequence industry that makes extensive use of modelling and simulation, and NASA formalised their approach to incorporating it into their activities in a standard, arising from the outcome of the investigation into the Columbia accident (NASA, 2016). It has an emphasis on applying V&V and UQ with a graded approach, by considering the consequence of a decision, and the extent of influence that M&S has on it.
- The development of safety cases requires processes to assess the credibility of evidence, much of which originates from modelling and simulation, and incorporating it into a structured argument. The US NRC have created a generic safety case in the form of a credibility assessment framework for models of critical boiling transition (US NRC, 2019b). They demonstrate the assessment of the evidence and maturity of component parts of the models by applying Goal Structuring Notation (GSN) – a method of expressing the structure and interrelationship of evidence and arguments, which is used to construct safety cases in a range of industries.

3.1.4 Examples of PIRT and PCMM

When applying PIRT and PCMM, there is not a ‘one size fits all’ approach, and their benefits only emerge through systematic use. This includes updating them at various stages of a programme, so that they can help to determine the most important or efficient steps to take next. Their utility is not realised if they are generated once, as an end in themselves at the start of a programme, and not revisited and revised.

The most instructive guidance that can be given in their application is to study some of the examples that have been published. All of the references cited below show specific examples of the application PIRT or PCMM (or both) to situations covering a broad range of maturity of both modelling methods and the reactor technology under consideration.

- It is possible to apply the most detail and specificity, in terms of phenomena, faults and accidents, predictions and validation comparisons, to reactors that are already operating.

These range from the original application to Loss-Of-Coolant Accidents (LOCAs) (US NRC, 1989, Wilson and Boyack, 1998) to re-assessment of the safety features of reactors (Wright *et al.*, 2014) and fuel storage (CSNI, 2018) following the Fukushima Daiichi accident. Similarly, detailed knowledge of plant behaviour can be applied to the process of creating and substantiating confidence in advanced multiphysics tools (Jones *et al.*, 2018).

- There are also examples for reactor designs that are not in existence, but because there is experience with existing reactors using the same technology, significant effort can be applied to planning their design and the strategy for their licensing using PIRT. The NGNP VHTR concept underwent such an assessment (ORNL, 2008a, ORNL, 2008b) along with its TRISO fuel (US NRC, 2004).
- Where reactor technologies are significantly less developed, the bounding faults and accidents are not necessarily known, and the modelling methods are not as mature. Here the PIRT and PCMM process can identify and prioritise fundamental research, modelling method development, and identify validation programmes that will be necessary for design and licensing. The US NRC used PCMM to assess the readiness of their codes for licensing all types of Gen IV reactors (US NRC, 2020). Some recent examples of where the PIRT process has been used to discover and combine areas of research and development for Molten Salt Reactors (MSRs) can be found in Diamond *et al.* (2018), Lin *et al.* (2019) and Singh *et al.* (2019), where the importance of UQ and SA in understanding the best areas to focus experimental effort is clear.

3.2 Approaches to Component Activities

Component activities are the second tier introduced in Section 1.1. The guidance discussed in this section is not intended to cover any particular type of assessment or modelling tool¹, instead the intention is to discuss the interactions between the difficulty of the tasks, the extent of available guidance and the needs and expectations of industrial analysis, particular in the context of the discussion in Section 1.4.

Table 3.1 is intended to clarify the nature of each component activity, the scope of which can often be ambiguous or misunderstood, if not clearly distinguished. Solution verification is defined here in relatively narrow terms, and broader activities are needed, so Quality Assurance Verification is also defined. Solution verification ought to fulfill a broader remit, but is often not discussed other than as relating to the details of numerical error (Roy, 2005), relating mainly to mesh resolution convergence.

¹ That information can be found in Sections 2.2, 3.3 and 3.4

	What is it?	How do you do it?	Why do you do it?	Gets confused with?	Why is it hard?
Code Verification	Assessment of errors in code solutions and order of convergence in response to mesh refinement.	Computing absolute errors for simple and exact analytical or manufactured solutions.	To confirm correct implementation of algorithms in code, and to identify algorithmic weaknesses.	SQE, benchmarking, code-to-code comparison.	Good reference solutions are hard to find; work is tedious and often unappreciated.
Solution Verification	Numerical estimation of errors in solutions for realistic problems.	Mesh refinement studies and error estimates for user chosen metrics.	To obtain estimates of the numerical uncertainty in quantities of interest for problems of interest.	Code verification, validation.	Problems of practical interest are often under-resolved and quantities of interest are often poorly behaved under mesh refinement; work is tedious and often unappreciated.
Quality Assurance Verification	Independent checking of all aspects of inputs, geometry, meshing, model selection and setup, solution, post-processing and reporting.	Detailed inspection, assessment or re-analysis by suitably experienced independent person.	Mistakes, oversights and unjustified assumptions in data gathering, modelling, interpretation and reporting of complex models are common.	Code or solution verification, validation.	Necessary to have skilled and experienced analysts who have not been part of the task to provide independence. Requires detailed record keeping and robust justification of choices.
Model Validation	Comparison of model results with experimental data or observations, or with higher-fidelity model results.	Quantitative comparison of experimental and model results, considering uncertainties in both.	To determine whether a model in a code is an adequate representation of reality for an intended application.	Verification, calibration, uncertainty quantification.	Experiments are often limited in quality, scope, or number, often have incomplete setup description and poorly understood errors, and do not cover enough of the application regime.
Calibration	Adjusting model parameters so that simulation results match data and observations.	Trial and error or optimisation and statistical approaches.	Sometimes a necessary activity to help ensure that a model produces acceptable results with some justification.	Validation, benchmarking.	Producing quantifiably 'good' results with inherently limited physical models is not easy. Users have complex problems that need answers, so they pragmatically 'do their best'.
Sensitivity Analysis	The systematic or ad-hoc study of how strongly input parameters or model choices affect outputs of interest.	Manual or statistical studies of inputs and their effect on outputs.	To discover and rank the importance of inputs, or to generate surrogate models.	Uncertainty quantification, calibration.	The large dimensionality of the input space limits the scope of formal sensitivity analysis; ad-hoc studies risk missing the importance of parameters and excluding them without good justification.
Uncertainty Quantification	Evaluation of how outputs vary with physically realistic variability in model inputs or originate in the applied methods.	Statistical studies of the distribution of outputs obtained by propagating distributions of input quantities.	Quantifying the potential range of values in the computed outputs, to compare available limits or margins.	Sensitivity analysis, validation, calibration.	Can be very resource intensive to examine the variability of all inputs and models. Realistic probability distributions or ranges of input quantities may not be known.

Table 3.1: Comparison of the nature and scope of component activities, adapted from Weirs and Winokur (2017).

3.2.1 Frameworks and Guidance for V&V and UQ

An introduction to the agreed definitions of the terms and overall aims of verification, validation and uncertainty quantification are given in Volume 1, Section 4.3. To extend from this, the intention here is to consider how they can be enacted practically, especially:

... where decisions are increasingly based on predictive simulations, V&V is about your personal reputation, your credibility as a scientist, and the consequences of your analyses.

Weirs and Winokur, 2017

This encapsulates the broad point that V&V and UQ and the process of gaining confidence in a prediction depends primarily on having an attitude and approach that does not assume that a numerical result is ‘right’ or adequate, but to take responsibility to make sure that it is sufficiently ‘right’ for its purpose, and for the correct reasons. This is a manifestation of safety culture, to have a *questioning attitude, a rigorous and prudent approach, and necessary communication* (IAEA, 1991).

This leads to another key point – it is not possible to simply use a piece of software that has been previously shown to be adequate or validated, and apply it in a different context, without giving thought to the reasons to believe it:

... the validation of a set of scientific simulation tools is case-specific. A tool cannot be said to be validated or qualified in general; however, it can be said to be validated for a given application in a given context.

Roelofs, 2019

Although even this point is open to challenge – to describe a tool to be ‘validated’ in a particular application and context, cannot be an absolute statement, all that it can mean is that sufficient effort and process has been applied so that there is no evidence that demonstrates that the predictions made² are unacceptably inaccurate for the intended purpose. What process to follow, how much evidence needs to be assessed and what level of disagreement would be needed to ‘invalidate’ the prediction is a matter to be negotiated between the analyst and the user (or assessor, approver, regulator) of the analysis, and potentially corroborated by peer review. A common view on agreed frameworks, standards and precedent (in the form of benchmark studies) provide mechanisms to facilitate this.

There are several references that provide a detailed elucidation of the basic terms and conceptual frameworks. The guidance provided by Oberkampf and Roy (2010) (alternatively in Roy and Oberkampf, 2011) is applicable to scientific computing generally. The ASME (2009, V&V 20) standard is applicable to V&V for CFD and heat transfer specifically, and there are widely cited references by Oberkampf and Trucano (2002), Roache (1997) and Stern *et al.* (2001) that also deal with V&V and UQ for CFD.

The topic should not be considered to be a settled issue, however, particularly given the gap that exists

² This is similar to the distinction between mathematics and science – in mathematics, ‘proof’ is possible; in science, a theory is accepted while it has power to explain and has not been falsified or contradicted by evidence.

between these formal and mathematically sophisticated concepts³, and the limited extent of their use in common practice (Roache, 2016). In the broader literature available, much of the guidance on V&V and UQ actually refers back to or is developed from the underlying ideas described by the authors noted above⁴ and ASME, so there is less plurality and independence in the elaboration of concepts than might appear. A new standard, ASME V&V 30, is in development that specifically deals with 'Verification and Validation in Computational Simulation of Nuclear System Thermal Fluids Behavior'⁵.

It will rarely be necessary in an industrial context to start a modelling task with a 'free hand' or 'blank page' in these regards because, particularly for QA, nuclear operators, developers and suppliers have established quality assurance processes and procedures. These encapsulate many error traps and provide significant opportunities for good practice, rigour and traceability, and will need to be adhered to when working in these domains regardless of other guidance.

3.2.2 Obtaining Validation Data

For large-scale model development and simulation supported licensing programmes, it should be expected that design-specific validation experiments will be necessary. Ideally a hierarchical method would be employed for complex problems (Oberkampf and Trucano, 2003), where evidence is established in the individual aspects of the simulation through validation experiments designed specifically for the purpose.

A new or emergent modelling problem will often not have any, or sufficiently complete validation data available, especially if it has not been simulated by an organisation previously. There is rarely time, budget, facilities or expertise to perform experiments to support a one-off or initial analysis. This means that the only option is to look in open literature for the closest applicable experimental case. Where suitable data can be found, it can often be too simple, hard to relate to the flow or heat transfer regime of interest and has an incomplete description for replication. For analyses that will be part of a larger programme of assessment, specific IETs may be commissioned, but even in those cases, it can be hard to obtain 'CFD grade' data at prototypical conditions. Further discussion can be found in Volume 1, Section 4.6. Modelling is also used to justify building (or not building) such large test facilities, or to help design them, and so the problem of insufficient validation data occurs at this stage too, where it is needed to add confidence to the decisions.

The suitability for validation purposes of papers that are found in the literature is variable, and their shortcomings for this purpose can only emerge after significant effort has been applied to replicating them. Experiments that are designed and reported with code validation in mind are substantially preferable. As a bridge between development programme-specific experimental facilities, and the open literature, benchmarking exercises and databases of simple reference cases exist to test and disseminate the capabilities of methods and tools, and are very valuable for validation. The available resources and past benchmarking activities depends on the type of modelling tool, and specific examples are discussed in Sections 3.3 and 3.4 for system codes and CFD respectively.

³ These authors are also not always in agreement (Oberkampf, 2002), so it may be necessary to 'pick a side'.

⁴ and their co-authors, primarily in US nuclear weapons community, although there is substantial and well-established cross-over to nuclear energy.

⁵ Along with ASME V&V 40 for medical devices and others in development for advanced manufacturing, energy systems and machine learning.

The US NE-KAMS (Nuclear Energy – Knowledge Base for Advanced Modeling and Simulation) project was initiated to build a validation database to assemble, interrelate, preserve and exploit existing multiphysics and multilevel (in a hierarchy) validation knowledge (details of experimental configurations as well as the resulting data) from wide range of sources. Its ambitious objectives (Weirs *et al.*, 2011) reflect the aims of the US nuclear V&V community (Oberkampf and Trucano, 2007). This project would be a highly valuable resource, and a prototype (based on the Gen IV Materials Handbook hosted by Oak Ridge National Laboratory, Ren, 2013) was under construction in 2017 (Ren, 2016), but more recent information is not readily available in the public domain regarding its progress or status.

An alternative approach to producing data is to apply higher fidelity models either to calibrate or supply inputs for lower fidelity codes and also to act as a surrogate for validation data. Examples are the derivation or assessment of the pressure drop characteristics for a system code component using CFD, or the use of Large Eddy Simulation (LES) or Direct Numerical Simulation (DNS) for comparison to or improvements of RANS CFD models (Merzari *et al.*, 2020). The use of higher fidelity modelling results as true validation is not universally accepted, but it is another possible item in a modelling engineer's toolbox.

3.2.3 Making Comparisons Between Validation Data and Simulations

Once suitable validation data has been identified, and simulation results intended to replicate it have been created, then a comparison is needed between them. This is not as straightforward a task as it might appear.

The most obvious initial approach taken is to plot the experimental data against the simulation results. In many cases this approach is the only comparison that is performed, and the assessment of the level of agreement and confidence provided is judged based on visual inspection alone, without the use of quantitative analysis. This can be a suitable comparison, if the intention of the validation is for exploratory (screening or initial decision making) purposes, aiming to determine if a simulation can qualitatively reproduce certain features and trends for a phenomenon of interest.

Visual comparison between validation data and simulation is, however, often disparagingly referred to as the 'viewgraph norm', because of its lack of numerical assessment. This criticism is heard strongly from those engaged in code verification. In that domain, simulation results can and should be assessed by comparison to analytical solutions and by the Method of Manufactured Solutions (MMS). These should produce numerical simulation results that are nearly identical to the comparison function. The very small remaining differences constitute rigorous error estimates, so how they behave under mesh refinement and depend on numerical algorithm implementation and convergence requires quantitative assessment. However, in validation, such precise quantification is not readily achievable.

Two additional features are needed to strengthen and make a validation data comparison more robust:

Uncertainty assessments: experimental data and simulations inevitably will not be exactly overlaid when plotted, but if the uncertainty in both were to be plotted as error bars, then the extent of overlap between the estimated error bounds can be evaluated. The simulation results would

need to be subject to UQ process that reflected the model and numerical uncertainty, and that also propagated as realistic an assessment as possible of the uncertainty in the geometry and conditions of the experiment. This means that the requirements for characterising the uncertainty both the setup and results of the experiment can be onerous.

Quantitative validation metrics: the difference between the values of the experimental data and simulations can be calculated as an error or difference term, potentially using a statistical approach where the PDFs of the uncertainty can be assessed. If a comparison is only being made at a small number of locations, then each can be assessed individually. If, however there is 1D, 2D or 3D field information, or transiently varying data, then a single evaluation, such as a sum of squares error may be appropriate. If the locations or times are not aligned between the experiment and simulation, then interpolation will be required. Measurements may also be clustered with irregular density (or the results in different regions are of varying relative importance) and so it may not be justified to assign a uniform weighting of each point. Therefore quantifying the level of agreement may contain substantial difficulties and choices in approach. Oberkampf and Barone (2006) review methods of defining metrics, although the topic is also not necessarily a settled issue.

Including uncertainty in comparisons presents an opportunity for perverse or paradoxical motivations. If it is judged to be a positive result for validation that the uncertainty bands overlap between experiment and simulation, then measures that increase the apparent uncertainties in either or both makes this agreement more likely (Coleman and Stern, 1998), which is at odds with a motivation to improve accuracy.

Comparisons of errors and developing metrics is valuable in activities beyond validation, for example, when comparing modelling options (like a choice of turbulence modelling approach) or the effect of mesh refinement. Synthesising the performance and agreement into a single value that can be ranked or plotted to compare simulations to each other has similar benefits.

Metrics are a particularly powerful tool if a specific tolerance is required, or when the metric itself has meaning that can be used to produce a more objective assessment of the adequacy of a prediction, and fair means of comparison. However, unless preventative steps are taken, it should be recognised that validation assessment can be subject to cognitive biases (such as anchoring or confirmation bias, Kahneman, 2011). It is not uncommon for the originator of a simulation, who is likely to be invested in its success, to be the one to compare the results to experimental data. The selection of which locations to compare, and how to compare them, as well as which modelling options to use, is subject to the tendency to keep reassessing and tailoring the evaluation until good agreement is reached, and to stop ‘digging’ into the detail at this point. It may need a conscious level of self-discipline, or assessment by an independent third party, to look closely and give sufficient prominence to areas of disagreement. Blind comparisons can avoid these problems, but in the case where the simulation is being performed for exploratory reasons, defining the locations, nature, purpose and tolerance of the appropriate comparisons in advance may be also be misleading.

3.2.4 How and When to Use UQ and SA

Adding to the confidence that can be placed in modelling predictions is an iterative process, which, for anything other than one-off or single discipline analysis, would benefit from being controlled, documented and communicated within an organisational approach. UQ and SA have a role in this, but for practical models in an industrial setting, how and when they are best applied in a specific organisation needs to be seen in the context of a typical model development process:

- A simplified sub-model or ‘toy’ replica of the problem containing the essential functionality that will be employed in the modelling software is often a valuable initial step. It provides a tractable ‘sandbox’ to test and practice with boundary conditions, material models, modelling options, naming conventions, mesh/nodalisation approaches and resolutions. It can be created with a minimal cell/node count, so can be debugged and optimised with rapid solution times. Many of the implementation problems that occur in industrial modelling are caused by working at scale – models with dozens or hundreds of separate components or boundary conditions to be defined rarely work first time. The knowledge and confidence gained because the intended setup works well on the ‘toy’ allows the problem solving approach to separate the task looking for the error that is stopping a solution from working, from the concern that the approach may not be the correct one, and may never work. It is common when trying to debug a large scale model to change many settings in a short period to attempt to find the source of a problem. It is easy to then forget to reset some of them once it has been resolved, so a minimally invasive and focused approach that the experience from the ‘toy’ can provide is valuable in preventing errors and confusion. This is well aligned with the use of pilot systems in software engineering, where it is usually beneficial to *‘plan to throw one away; you will, anyhow’* (Brooks, 1995)
- Sensitivity analysis (using the sampling and propagation methods that are common to it and UQ) can provide numerical results and can be applied at any point in an analysis process, but only gives insight to the parameters selected for analysis. Real models contain too many inputs to address the effect of each on the results. An initial sensitivity assessment approach of screening them will be essential. The aim is to decide what matters – to quickly determine how well-known parameters are, and if uncertainties in them are significant in the output metric. This usually involves ad-hoc and exploratory work and should emphasise the identification of whether expected phenomena can be reproduced. Despite being exploratory, omissions or errors in this process can be significant if they mis-characterise the importance of an aspect, so documenting the process and decisions, producing meaningful metrics (similar to or the same as those discussed above for validation) and making it readily repeatable are advantageous.
- During the various stages of constructing and using a model, the plant design and operating conditions may evolve, and more refined information will be available from other modelling or design disciplines, and so initial screening assessments may need to be revisited, and updated ‘baseline’ results produced that reflect the new condition. Automation of solution processes and consistent post-processing of key components is very valuable to allow the effect of these changes on the output metrics to be consistently evaluated.
- Despite well developed methods and frameworks being available, attempting to apply formal UQ and SA methods rigorously too early is likely to add expense and confusion to real

models of complex industrial problems. Such a model needs to achieve a sufficient level of maturity and robustness first. While a model remains under-resolved, difficult to converge and its predictions still sufficiently approximate when compared to experimental data, then these methods cannot provide insight with the certainty they promise, and the effort is not justified. It may be the case that models never become good enough to be suitable to apply them, or the answers that the model produces are sufficient without applying them.

- UQ and SA are valuable activities to perform when a model has been developed to the point where its outputs are going to be used in a decision or substantiation. UQ allows better validation comparisons to be made, BEPU arguments to be made for safety assessment, and the size of a margin to be quantified in a conservative calculation if challenged. SA, by being able to quantitatively relate how changes in inputs affect output metrics, illustrate how well these features need be controlled to maintain a desired margin or performance. Guidance on how to interpret the outputs of UQ and SA is given in a number of places in this technical volume:

Section 2.2.3: Visualisation.

Section 4.5: Sensitivity Analysis.

Section 4.7: Assessment of Output Uncertainty from Propagation.

Section 4.8.5: Presenting Overall Uncertainty in an Output Metric.

This description does not indicate where QA verification should be performed, the level of code verification necessary (which will depend on the maturity of the software), and whether a focus of effort on solution verification is warranted.

The integration of validation comparisons also depends on whether the comparisons are being performed with the full model, or using existing literature data from a related, but separate configuration. In the latter case, the learning available from performing these validation simulations early in the process will aid the model development. This can be done alongside the 'toy' model, or in some cases it can also serve as the 'toy'. If experimental results representing the full model configuration are being produced while model is being constructed⁶, then iterative loops of experiment, simulation and comparison will be required, and the amount of UQ necessary at each will need to be chosen to suit.

3.3 Methods for System Codes

System code methods evolved during an extended period of reactor design and construction, and with the benefit of large scale experimental facilities able to produce modelling closures and validation data. They are the accepted and underpinning technology for reactor analysis and licensing, and as such there are mature approaches to producing analysis which commands confidence, and there are established UQ methods, all of which are well documented. This is not the case with CFD, which will be discussed separately (Section 3.4).

This history, and their role in the development of BEPU methodology means that the available resources for discussing them is extensive, and this section intends mainly to highlight overarching

⁶ and may be required to provide experimental inputs to calibrate hard-to-characterise features.

principles, large scale programmes and example benchmark activities associated with confidence and uncertainty. The following books, reports and papers are suggested as a starting point:

Volume 1, Section 4.6: The CSNI validation matrices for Light Water Reactors (LWRs) are discussed, as are the basics of scaling.

US NRC (2005) and Wilson (2013): The established nature of demonstrating the adequacy of system code models is well described, as is the motivation and relevance of understanding uncertainty, and links to the organisational approaches described in Section 3.1.1.

IAEA (2002), IAEA (2003) and IAEA (2008): Provide guidance and details for when to use conservative and BEPU methods. To avoid potential confusion between the terms ‘deterministic’ and ‘probabilistic’ analysis in these references, it should be noted that the probabilistic aspect refers to the likelihood of occurrence of initiating events for accidents or transients and availability of plant systems, and a deterministic analysis models a particular set of conditions. A deterministic simulation may be conservative or best estimate in nature, and will still make use of uncertainty quantification methods (IAEA, 2019).

D’Auria (2017): Chapter 13 discusses V&V, uncertainty and scaling for system codes, and Chapter 14 covers how BEPU can be applied.

Martin and Frepoli (2019): Scaling, V&V and best estimate methods are covered.

A discussion of the sources of uncertainty in system code thermal hydraulic calculations is given in IAEA (2008, Annex I). A key point, particularly in contrast with CFD is the extensive use of and need for simplified models and closures/correlations in system codes. They take account of complex physics, reduce computational expense, enable high level modelling and are the manifestation of where experimental data is captured and exploited. The validity range and implementation details of these correlations, as well as uncertainty in their source data, can have a significant impact on the uncertainty introduced into the overall calculation. The use of correlations is relevant to subchannel codes too (Salko *et al.*, 2016, Jones *et al.*, 2018) although these are not discussed in detail here.

Linked to the use of correlations is the issue of scaling. Correlations are typically based on non-dimensional characterisations of the particular mass, momentum or energy transport process they replicate, and have a range of validity based on how they were derived. Reduced scale test facilities inevitably lead to distortions of the replication of the geometrical aspect ratios, flow regimes and timescales (Dzodzo, 2019), because, compared to a full-sized reactor, they can only cover a limited range of these non-dimensional parameters, and cannot align them all simultaneously. This leads to errors, uncertainties and limited validation coverage in the system code models which are derived from or validated against such scaled tests. The links between scaling and uncertainty are discussed in CSNI (2017b) where common uncertainty methods are described, based on either the propagation of input uncertainty (as covered in Section 2, such as CSAU or the GRS method, Glaeser, 2008) or methods based on accuracy extrapolation (UMAE). These three methods receive prominence in much of the international literature, partly because there is commonality in the contributing institutions and authors that developed them for several publications, so, like with V&V frameworks, the implied level of consensus, and the extent of coverage of a diversity of approaches⁷ may reflect this.

⁷ or evaluation of their consistency with each other or completeness (Pourgol-Mohamad *et al.*, 2011)

Validation is also linked to scaling, where the effect of scaling distortions can be mitigated by using experimental data from facilities of different sizes and based on different scaling approaches – these are known as a counterpart tests (CSNI, 2017b).

The OECD NEA CSNI have organised an ongoing series of BEPU thermal hydraulics modelling programmes over more than 25 years:

- Uncertainty Method Study (UMS).
- Best Estimate Methods Uncertainty and Sensitivity Evaluation (BEMUSE).
- Post-BEMUSE Reflood Model Input Uncertainty Methods (PREMIUM).
- Systematic Approach for Input Uncertainty quantification Methodology (SAPIUM).

The completed phases are described in CSNI (1998), CSNI (2011a) and CSNI (2017a) (or Nouy and de Crécy, 2017, Skorek *et al.*, 2019). The lessons in the first phases led to PREMIUM giving increased attention to input uncertainty, and is currently being expanded by SAPIUM to use IUQ (Baccou *et al.*, 2020), where the uncertainty in inputs is assessed based on the comparisons between experimental data and model predictions. It should be noted that in these studies, user effects were observed both in the creation of model predictions, and the application of UQ methods.

The preceding discussion is focused on the application of system codes to LWRs, where system code closures were developed from extensive experimentation and the UQ methods for them can take advantage of this experimental underpinning. For new designs of Gen IV reactors, there is less experimental evidence, particularly from IETs, although the underlying principles of modelling and UQ will be largely applicable. For example, international benchmarking programmes have been conducted where a range of organisations modelled tests conducted on two historical Sodium-cooled Fast Reactors (SFRs): Phénix (IAEA, 2013) and EBR-II (IAEA, 2017). The same EBR-II tests were recently modelled using SAM, with Dakota using Latin Hypercube Sampling (LHS) to sample inputs to apply UQ to the prediction (Mui *et al.*, 2019)

3.4 Methods for CFD

The application of CFD to nuclear safety is a topic that is under active consideration by international organisations and national regulators. This scrutiny places a requirement to account for and demonstrate uncertainties within a CFD analysis with a high level of confidence, with a focus on validation comparisons, on the ability of the chosen approach to turbulence to model the phenomena encountered, and on demonstrating the adequacy of mesh quality and resolution (Downing *et al.*, 2018).

The experience and methods originally developed for system codes in NTH simulations described above have been used as some of the starting points for applying UQ to CFD:

- In addition to publishing CFD best practice guidance (CSNI, 2015) the CSNI have presented a CFD specific summary of sources of uncertainties, have contrasted the characteristics of uncertainties for (single-phase) CFD with system codes, and provided some examples of the application of a range of different methods (CSNI, 2016b). This document serves to highlight that UQ in the context of CFD for NTH is currently not a settled issue, because there are few

actionable recommendations and the description of the methodologies is not amenable to being adopted by readers (reinforcing the position described in Section 1.4).

- CSAU is a UQ method applied to system codes, and Acton and Baglietto (2019) have considered what would be needed to apply it to CFD, also providing a summary of the needs and differences of the use of CFD compared to system code justifications. Uchida and Kawamura (2008) also describe error estimation for CFD simulations in the CSAU framework.

In its broader industrial application, however, CFD has paid attention to V&V and UQ for several decades, driven in many cases by aerodynamics (Walters and Huyse, 2002); some of the CFD specific frameworks described in Section 3.2.1 originate in the aerospace CFD domain. This is understandable, given that, while the prediction of lift, drag and separation on air-frames is a challenging task, it is possible to make very detailed measurements, and tune turbulence models to suit. It does not have the range of length scales that can be encountered in NPP simulations, nor needs to predict, for example, buoyant heat transfer or two-phase flow.

The propagation of uncertain inputs through a CFD model can be achieved using the same methods as any other calculation, however, CFD model evaluations are almost inevitably computationally expensive, and so any UQ or SA method applied must anticipate the fact that the number of runs will usually be limited to dozens (or hundreds), and so the most benefit must be gained from the chosen set. This places emphasis on the optimisation of mesh, modelling choices and early sensitivity assessment to select the best setup.

3.4.1 Model and Numerical Uncertainty in CFD

Parametric and initial/boundary condition uncertainty only accounts for some of the overall uncertainty in a prediction. There are a number of other features that are notable about uncertainties and UQ for CFD that create model and numerical uncertainty:

- The topic of mesh independence and mesh resolution has been extensively studied, and rigorous methods such as Richardson extrapolation or Grid Convergence Index (GCI) (Roache, 1997, ASME, 2009, V&V 20) are able to quantify its effects. For transient simulations, it is equally important to look into time resolution, although the Courant-Friedrichs Lewy (CFL) number provides a good guide, and the timestep cannot be varied independently of mesh resolution (smaller cells need smaller timesteps to maintain a given CFL). To establish an appropriate time step, the timescales or frequency of the modelled phenomena should also be considered, as discussed in Volume 3, Section 3.2.
- The approach to modelling turbulence, in terms of using RANS (or URANS), LES or hybrid methods, and which model and options to choose within these types will often have a determining effect on the accuracy of a simulation.
- The numerical uncertainties introduced by the discretisation of the PDEs of the Navier Stokes equations via the finite element or volume method, and the discretisation of spatial and temporal gradients⁸ can in some situations be estimated, but for complex 3D flows in unstructured meshes this is very challenging and not routinely assessed.

⁸ Often described as using 'first order' or 'second order' methods, although this simple distinction belies a large range of possible options and their characteristics.

- The set of transport equations in CFD codes are always solved by some form of iterative algorithm. In ideal cases, the residuals of these equations can be decreased unambiguously to steady values below a specified tolerance. However, in the case of complex flows or more complex physics (like combustion), where mesh resolution or quality is inadequate, or where there is inherent unsteadiness in flow fields, then the residuals (and other output metrics used to monitor convergence) are often not stable, and cannot be reduced without actions that introduce more uncertainty⁹.
- Some uncertainties are specific and relevant to only some problems, for example:
 - The interaction of buoyancy with turbulence.
 - The prediction of shockwaves and sonic/choked flow.
 - Porous model simulations, particularly the adequacy of the porous representation, its interaction with adjacent regions of non porous fluid, or the treatment of turbulence within the porous region.
 - In resolved turbulence simulations (such as LES), correctly preserving the (spatial and temporal) frequency content can be highly important in some applications (although is less important if only the mean flow fields are needed).
 - In multiphase simulations, the treatment of free surface motion has several options and can introduce significant uncertainties, especially in situations where surface tension is important.
 - When modelling thermal radiation, the chosen model and numerical settings can have a strong influence on the cost and accuracy of a simulation.

Discretisation schemes and turbulence models are 'categorical variables', because they represent discrete choices, and there is no inherent ranking of their characteristics, so they cannot be propagated through a UQ method using an approach that samples a probability distribution. However it is possible to include them, as well as the numerical and mesh resolution uncertainties, if they can be estimated, using the methods discussed in Section 4.8 as epistemic uncertainties. The extent that this can be expected to be achieved for some of these topics is discussed in Section 3.5.2.

3.4.2 Validation Data and Benchmarks

High-fidelity Validation data that is relevant to a new problem to be modelled by CFD can be difficult to obtain. There are references to databases for data in Volume 1, Section 4.3.2 and reviewed in more detail by Frazer-Nash (2019, Section 7), which are particularly related to relatively simple flows (basic tests and SETs).

When time and cost constraints are applied in an industrial context, the question will inevitably arise: *Do I need to validate this? Can I trust that the CFD code will be good enough for this problem?* The answer to this question depends on the problem at hand, and the consequences of using or placing trust in the answers, but it is possible to use the experiences reported from international benchmarking activities to guide this.

The IAEA have published studies relating to reactor and fuel analysis (IAEA, 2014, IAEA, 2020a, IAEA, 2020b), and a longer and broader series of benchmarks have been organised by the CSNI.

⁹ For example using first order 'upwind' differencing for the convection term in the momentum equation damps oscillatory processes, but introducing this diffusivity makes the velocity solution less accurate.

In the benchmark studies a number of groups each simulate a well characterised laboratory experiment that has high-fidelity data available, using methods of their choice. The topics covered are:

- Mixing at a T-junction, giving rise to thermal fatigue (CSNI, 2011b).
- Turbulence downstream of a rod bundle spacer grid (CSNI, 2013).
- Mixing of a buoyant jet with a stratified layer in a large volume (CSNI, 2016a).
- Turbulent mixing of adjacent streams with a density difference, including the application of UQ (CSNI, 2019).
- The buoyancy driven mixing of fluid in a scaled representation of the cold leg and downcomer of a PWR (Orea *et al.*, 2020, CSNI report not yet released at time of writing).
- FSI of two inline cylinders, predicting the flow field and structural response (in progress at time of writing).

A key feature of the exercises is that they have an ‘open’ and a ‘blind’ phase. In the open phase, which happens first, the experimental setup and results are provided and the quality of the prediction agreement can be judged by each group, and the modelling approach optimised. The blind phase follows this, where the same experimental apparatus is run with a difference in conditions. Only the setup is provided to participants and their predictions are submitted without being able to compare them to the experiment. This replicates some of the features of trying to make predictions in real simulations, because it tests how transferable the conclusions reached in the open phase, where validation data is available, to a similar condition where it is not.

An objective of these benchmarks is to compare the predictions made by a range of methods, meshes and codes, that are chosen by the participants. This allows an assessment to be made of user effects (for example, choices relating to the location and characterisation of flow inlets) and comparisons to be drawn¹⁰ between CFD codes and turbulence modelling approaches, and some guidance on best practices to be developed for the item of plant at-hand. A notable feature of these benchmarks is, however, the variability in the predictions made in some of the cases. This is even the case where an effort has been made to define a ‘common model’ for some of the results (Andreani *et al.*, 2019).

While they generate publicly available data, detailed official reports and additional publications from the participants (Cutrono Rakhimov *et al.*, 2019, for example) as well as international collaboration and cooperation, these exercises take several years to complete and report. They also only address specific items of plant and phenomena. Therefore, while they serve a highly valuable purpose, if the problem to be modelled at hand is not covered by them, then their outputs and lessons can only be used as a starting point for asking questions, rather than providing readily applicable answers.

Another conclusion that can be drawn from these benchmarks is that accumulating experience with a type of plant and its phenomena, and evaluating suitable meshing and modelling options is very important. However, relying solely on this experience when applying CFD to a new case has risks – there is still a need to exercise care and re-evaluate mesh and modelling decisions.

¹⁰ or at least the dependency of predictive performance on these parameters is highlighted – often the various benchmark results have too many differences to draw firm conclusions.

Part of the difficulty in making transferable conclusions is that the comparison of the effects of differences in setup and results is not fine grained enough. There are too many physical processes occurring and too many modelling differences present at once to separate out what really makes the differences. For example it is noted in Andreani *et al.* (2019) that unless the representation of the spreading of a jet as it enters the domain is correctly predicted (which can be assessed separately), the rest of the model prediction of that jet's behaviour will be compromised. Studies A and D also provide examples of some of the challenges in comparing CFD models to validation data (where, in both cases, the experiments were designed specifically for CFD validation), and unravelling the contributions of and sensitivity to various physical and modelling aspects.

Unless the the component parts of a simulation can be assessed and optimised, then the assembly of the validation arguments for the whole simulation may be compromised. This is aligned with the discussion in Section 3.2.4, where the most important stage of an analysis is 'getting the basics right', and knowing how costly or practical this is. It may be the case that adequately resolving each feature may be too costly. In the past, when computing hardware was a more significant limitation, then this was certainly true (Dzodzo, 2021).

The conclusion from considering how benchmarks affect how to answer the question: *Do I need to validate this?* is to dismantle the whole problem into its parts and to reverse the question: *what are my reasons to believe that this part is adequate?* By doing this, a hierarchical approach can be assembled, whereby a larger and more complex problem with many features and phenomena can be assembled and used for predictive purposes¹¹, with some confidence that none of its components is going to contribute unbounded inaccuracies. System codes have been through this process over decades, and have gained acceptance despite containing substantial empirical elements, but CFD codes are still not as widely accepted for safety assessments. Applying UQ to simpler parts of more complex model hierarchy, and demonstrating that they can be used to establish the uncertainty in a realistic model may be the only method that can provide sufficient confidence to allow BEPU CFD simulations with safety significance to be accepted.

3.4.2.1 CFD used for Safety Justification: Dry Spent Fuel Storage

It is instructive to recognise that there are areas with nuclear safety implications where design and licensing decisions have been made with a significant contribution from CFD. The dry storage of spent LWR fuel in pressurised canisters requires the Peak Cladding Temperature (PCT) of fuel elements at the centre of these canisters to be predicted with confidence, so that it is assured that they will not fail in handling or storage. This is an example of the type of output metric discussed in Section 2.1.3 where the PCT is the hottest part of the hottest fuel assembly (hence is in the tail of a probability distribution) which will contain modelling uncertainty, and the probability of failure of a fuel pin will also depend on its internal pressure, which will be represented as a probability distribution of its burnup history, which will contain epistemic and aleatoric uncertainty.

This is a difficult modelling task for lower fidelity models, like system codes, because it involves 3D Conjugate Heat Transfer (CHT), radiation heat transfer and buoyancy driven flows, where subtle pressure loss and driving force balances need to be established. However it is a tractable task for CFD because the equipment is not very large or complex, and so does not need a prohibitive

¹¹ This must be the objective for any modelling activity, otherwise if complete and prototypical validation is needed for each simulation, then modelling would serve little purpose, and the test results would simply be used.

number of cells. The fuel bundle can also be represented with a porous medium model, radiation and conduction can be included in most CFD codes, and the buoyancy and pressure loss can be modelled without substantial approximation.

This has led to dry spent fuel storage being a ready-made example, with documents published describing specific CFD best practice for passive cooling and CHT, plus consideration of validation and uncertainty (US NRC, 2013, US NRC, 2019a, Li and Liu, 2016). This also demonstrates the point that NTH does not need to be restricted to analysis of in-reactor primary circuit components, and is required for a range of applications.

3.4.3 Applying Polynomial Chaos

Polynomial Chaos Expansion (PCE) is one of the UQ tools most often applied to CFD in recent literature, relating to, for example, aerodynamics (Hosder and Walters, 2007, Schaefer *et al.*, 2016) as well as its application to NTH (CSNI, 2016b, Khalil *et al.*, 2018, Wenig *et al.*, 2021). The concepts of PCE are introduced in Section 4.4.2.4.

PCE can be relatively easily applied to create surrogate models for a small number of output metrics from a simulation, such as maximum values, combined/integrated values or values at a specific location. This provides the mean and standard deviation of the output metric, the Sobol indices for the influence of the uncertain parameters, and the surrogate models can be further sampled to determine the CDF of the outputs. The uncertain parameters chosen to vary, and the number and distribution of CFD solutions required will need to be chosen carefully, particularly if they are expensive to evaluate. Study B provides an example of this.

It is also possible to apply PCE to CFD results more broadly, either determining the uncertainty and sensitivity at all points along particular sampling lines across the domain, or at every location on the surface or in the volume of a mesh. This allows the uncertainty (or sensitivity) in the velocity, pressure, or temperature fields (for example) to be visualised using the normal methods of post-processing. There are ‘intrusive’ methods that achieve this, where the PCE is implemented within and by modifying the underlying equations in the CFD source code, and only one solution is needed of the revised equation set (Najm, 2009). However, the source code may not be available, and the modifications are likely to be difficult and time consuming, particularly for problems with complex physics, so are unlikely to be applied by industrial users. There are also ‘non-intrusive’ methods where the CFD code is used without modification, and a sampled set of solutions are performed; a PCE model is fitted at each location of choice (which may be all locations in the mesh).

Developing a PCE result to allow the uncertainty at any point in a flow field should work well for steady-state flows or unstable or transient flows with well converged time-averaged fields. However, a flow may contain sudden switches in flow regime, such as large changes in the presence, location or size of separation, multiple possible stable states (for example a jet attaching to either one wall or another of an expanding duct), or laminar-turbulent transition over the range of uncertain parameters studied. In this case, interpreting the ‘uncertainty’ simply from the statistics may be misleading, or give an incomplete or substandard insight (the mean of a result with two distinct states will not represent the physical behaviour of either). It may be necessary to form sub-populations of solutions grouped by one of these distinguishing regimes, and then gain more understanding by studying uncertainty within and between groups, including finding the interface

(‘surface’ in parameter space) where the switches occur.

There are also no widely used commercial or open source CFD codes that currently incorporate full-field PCE functionality, and there are no mature and available interfaces to toolkits (such as Dakota) that are able to read the CFD results files for each evaluation, generate the PCE models, and then write data out in a suitable format for visualisation. Therefore, applying non-intrusive PCE currently requires an investment in capability development. This is not facilitated by available literature, which is dominated by descriptions that approach the topic from a mathematical perspective, rather than with a more approachable implementation and application focus.

3.4.4 Proper Orthogonal Decomposition and Machine Learning

Runchal (2020, CFD of the Future: Year 2025 and Beyond), reflecting on 50 years of CFD and its future prospects predicts that Machine Learning (ML) will impact many aspects of it: data reduction, model interpretation and post-processing, solution algorithms, as well as data driven turbulence models.

Notable types of surrogate model receiving prominence in CFD are Proper Orthogonal Decomposition (POD), as a form of model order reduction (Section 4.4.2.6), and those associated with ML – Kriging (Section 4.4.2.5) and neural networks. These are able to use CFD simulation results to derive or ‘train’ a surrogate model, which can then rapidly¹² emulate either the time evolution of the same flow system, or its response to parametric variation, or both together.

Like UQ, the available literature on POD is not generally aimed at practising engineers, however, the tutorial by Weiss (2019) serves as a straightforward entry point to its application, and the overview by Cordier and Bergmann (2008) provides more detail and discussion that is still approachable. Most CFD codes also do not provide this functionality natively, but the toolbox used by Georgaka *et al.* (2020) is open source¹³, and the functions needed to perform POD are matrix operations and the determination of eigenvalues and eigenvectors, which is functionality that is commonly available (in MATLAB or Python for example). Georgaka *et al.* (2020) describes the application of POD to thermal mixing at a T-junction, and Dolci and Arina (2016) provides a description of the application of POD methods for aerodynamics. The structure of flows in fuel assembly rod bundles and T-junctions have been studied with the assistance of POD (Merzari and Ninokata, 2011, Merzari *et al.*, 2013) and it can be equally well applied to experimental results of similar geometries – for example Nguyen *et al.* (2020) extracted the dominant flow structures in the presences of localised blockages in a rod bundle.

How to identify the dominant (most energetic) modes in turbulent flows, interpret their structure and emulate or reconstruct them is a focus of most of the attention for ML. Approaches are proposed that aim to use ML methods to improve turbulence models (Duraismy *et al.*, 2019) and there are methods that aim to avoid modelling turbulence in the conventional way altogether (Hanna *et al.*, 2020), replacing it entirely with data driven models. An example combining many of the aspects discussed are the methods demonstrated by Yeh *et al.* (2018) and Chang *et al.* (2019), where the data from LES solutions (selected using a DoE method) are reduced using POD and using Kriging

¹² Potentially three or four orders of magnitude faster than the original CFD.

¹³ mathlab.sissa.it/ithaca-fv

(aided by Sobol indices) to build a physics-based emulator of the flow in a rocket engine injector, to allow the design space to be surveyed.

It should be noted that, like UQ methods, the additional complexity and sophistication of POD and ML comes with additional costs – if they play a part in a simulation used to make a decision with significant consequence, then a proportionate level of effort is required to demonstrate that they are functioning correctly, to estimate the errors and uncertainties that they contain, and to understand where they reach their trustworthy emulation limits. This will be challenging for non-specialist users, and, because their internal operation can be hard to inspect, may prove prohibitively hard to demonstrate when modelling a high-consequence, safety critical scenario.

3.5 Applying Effort in the Right Place

When considering how and when to apply the methods discussed throughout this technical volume, it is essential to remember to not become embroiled in or enamoured with their mathematical and numerical details, at the expense of maintaining a clear link to reality for any V&V and UQ. Understanding the uncertainty rigorously and comprehensively on a relatively small model is tractable (including the sources of numerical and modelling uncertainty), but adding together input and material property uncertainty, as well as a range of potential plant conditions makes fully addressing UQ for all aspects simultaneously expensive, and risks making it hard to form a clear picture of the main effects and where to focus efforts. A progressive approach, building understanding piece-by-piece is likely to provide more insight, but cannot necessarily maintain completeness or rigour.

This section intends to provide a discussion about common error traps, misconceptions and omission of sources of uncertainty, as well as potentially misdirected efforts, which can arise if the links to reality and context of an analysis are not sufficiently borne in mind. It is also intended to provide some prompts to answer the last question posed at the start of Section 1 – *how much effort is warranted...?*

3.5.1 Change Control and the Validity of Geometry and Input

In a design, construction or licensing process, or during operation, changes will be made to plant geometry, operating conditions, the characterisation of transients, faults and hazards, or to safety thresholds and limits. These changes may invalidate previous NTH assessments, requiring them to be repeated. While this could be costly, it is less problematic or insidious than decisions being made using the results of assessments that were made using obsolete or invalid information. This could easily be a larger source of discrepancy than the other uncertainties in the prediction.

Preventing this error trap with confidence requires robust change management processes to ‘pull’ and ‘push’ the necessary updates:

- An analyst should ask themselves: *Am I using the latest geometry or conditions? How can I traceably define the state of a model's inputs, and compare them to the current source of 'truth'?*
- The significance of some changes that are made in the design or manufacturing stage for a component may be obvious to predict with respect to their effect on the NTH behaviour, or the

necessity for a specialist review may be covered by change management rules. However, it can be the case that changes to small features, such as to the presence or absence of chamfers at openings, or the introduction of contractions or geometry changes that lead to reduced open areas in flow paths may not be noted as requiring assessment. A similar point applies to anticipating the consequences of changes in operating conditions.

These are both potential sources of unsafe latent conditions – holes in the ‘Swiss cheese’ (Volume 1, Section 2.2.3). While it is an organisational (rather than a modelling) challenge to recognise, anticipate and prevent them, experience has shown that, for the latter type, the significance of some changes for plant flows and heat transfer can be missed, misunderstood or underestimated by engineers from other disciplines, who are not specialised in NTH.

3.5.2 Code and Solution Verification vs. Practical CFD Models

Significant attention is paid in the V&V literature to code verification, and to demonstrating the ‘convergence order’¹⁴ of a software tool. Code verification of the implementation of a method assesses its convergence order compared to the theoretical behaviour, either by comparison to analytical solutions, or by using MMS. Approaches are well advanced in a research context (Roache, 1997, Salari and Knupp, 2000) and defined by ASME (2009, V&V 20). The methods and techniques are also still being developed and expanded (Krueger *et al.*, 2019). Guidance on calculating error estimates and demonstrating grid convergence for solution verification is also well established, (e.g. using GCI, Section 3.4.1).

While important, much of the formal code and solution verification guidance is applicable largely to software developers. Ideally it would be made available by the developers of CFD codes, and potentially run by users as acceptance, and confidence building tests. However, most CFD users are not developers, and most real problems rarely possess the straightforward and benign problem characteristics that allow formal order of accuracy assessments and GCI to be pragmatic and informative. Practical models often need to have automated generation of complex unstructured meshes, with local refinements and mesh quality optimisations, and can exhibit turbulence model behaviour and large scale or local flow instabilities that spoil the idealisation. It is also the case that the expected formal convergence order can be mathematically proven for simple idealised geometries, but are unlikely to be for 3D meshes of cells that are skewed/non-orthogonal, or are complex polyhedral shapes, and apply more complex numerical discretisation methods.

Judging mesh resolution adequacy can be hard, and formally converging the mesh (via Richardson extrapolation, for example) is often impractical for mesh size reasons. Users will need to consider whether they could ever afford to properly converge a whole domain (especially for a transient, or when needing statistically meaningful and stable averages for an unstable flow). For a large or complex geometry, computing limitations or costs mean that it is likely to be only worth applying effort to considering the resolution in some key areas of interest.

Code verification is partly used to capture implementation errors (bugs) in numerical algorithms. Bugs in modelling tools can lead to significant errors in a range of areas, not just related to the

¹⁴ Numerical PDE solutions equations which are truncated via a Taylor-series expansion with a certain number of terms, for example, have an error term that depends on the size of a cell, δx or timestep δt , and the order, p , is the rate at which the error reduces with refinement, e.g. with δx^p or δt^p .

Methodologies

solution of the PDEs (for example in post-processing). Open source CFD tools exist, of varying quality and maturity, but bugs are also still present in established commercial/proprietary codes. The contents and quality of any code are a product of its history and application, and new or infrequently used functionality is riskier to rely on. Open source tools typically have different types of bug and failure mode compared to commercial codes, but the availability of the source code means that users can find and understand them. Error reporting mechanisms exist for both open source and commercial codes, and these should be available to and routinely scrutinised by users of the software. The error report will state the versions of the code affected by the error, the affected functionality and consequences of the problem, ways to work-around or avoid it, and when the problem was rectified.

This discussion also needs to be put in the context of considering whether the numerical and mesh errors make a significant contribution to the overall uncertainty or assessment of the results of a specific CFD model:

- If the uncertainty in inputs is large, then there is less point in being concerned about smaller uncertainties in the model. If the model and numerical uncertainties are dominant, then they do require attention.
- Likewise, if the margin available to a safety or performance limit is large, and the model behaviour is benign (there are no 'cliff-edges' in its predictions), then applying a significant effort is not beneficial.

These points should be borne in mind when considering the extent of solution verification necessary, as well as the fact that, in practice, it is difficult. A problem may need to be under-resolved for pragmatic reasons, and high quality validation data may not exist.

3.5.2.1 Unsteadiness in Steady-State CFD Solutions

If a flow is complex, for example with bluff bodies and separation, dividing and combining flow, or large buoyant plumes, then a CFD solution will always be inherently unstable or oscillatory to some extent. It is often the case that these applications are attempted to be solved with steady-state solutions, because the unsteady effects can be localised and creating a statistically converged transient result to obtain averaged results can be significantly more expensive. In the steady solution approach, it is typical to see variation in the quantity of interest – 'jumpy' behaviour of monitors of pressure, force, temperature etc, or poor residual convergence behaviour. It is tempting to try to avoid this behaviour, for example by using lower order discretisation or reducing under-relaxation factors, which can suppress dominant numerical instabilities, and allow a converged solution to be found. However, by taking this approach too far, it is also possible to artificially suppress the jumps by inhibiting the physically unstable features of the solution from progressing or varying.

When this unsteady behaviour is seen, then it can be beneficial to change the simulation to a transient as a comparison case, comparing the averaged flow fields or monitored values to their steady versions¹⁵. In some instances, running transiently makes no difference to the results; in

¹⁵ The concept of the time average of a fluctuating quantity being equal to the 'ensemble' average of many independent shorter realisations of this quantity is known as the quantity being 'ergodic'. Ergodicity has a precise meaning in the statistical description of turbulence, related to the Reynolds decomposition underlying the definition of RANS (Deissler, 1992). The description here relating to practical comparisons of steady and unsteady flows is too 'hand-wavy' to be considered a realisation or test of the ergodic nature of the flow – more care and rigour would be needed (Makarashvili *et al.*, 2017).

others it makes a substantial difference. Repeating this judgement in a quantifiable way with robust statistics to pursue a formal mesh refinement is unlikely to be tractable, and will rarely be practical in an industrial analysis context (with the computational resources currently typically available).

It is not possible to give general guidance on when these more complex cases occur. It should be considered on a case-by-case basis, and it is the responsibility of the analyst to look for and demonstrate that these issues are not present. In some cases, it will be necessary to solve URANS instead of steady-state RANS and time average. In some cases, a resolved turbulence approach, such as LES will be necessary. Either way, it should not be considered sufficient to simply focus on mesh refinement (global or local) to assess and assert that the solution is adequate.

3.5.3 Should UQ be Applied to Every Analysis?

This technical volume is intended to provide sufficient introduction and motivation to allow practising NTH engineers to think clearly about the confidence that is placed in their simulation results, and indicate the tools to determine and express the uncertainty in them. These final remarks are intended to provoke some consideration of how UQ will fit into their circumstance before committing to (or advocating that) it should be applied to everything:

- UQ does not ‘come for free’ – it can be expensive, and adds the additional burden of asking questions not only of the adequacy of a simulation, but also whether the UQ methods and outcomes are adequate. If it is going to be relied on to aid a decision or justification, it has to be high quality, good enough, trustworthy enough to justify that.
- For organisations and users of analysis results who are used to making assessments with deterministic analysis results, then applying a probabilistic uncertainty to the result can undermine confidence in the prediction, not add to it. It is understandable to feel that the precision of an answer produced by a long established ‘baseline’ best estimate model implies its accuracy and certainty. It is not necessarily welcome to be confronted with a (potentially large) uncertainty estimate, where there was none before. The stakeholders in an analysis may need to have the implications and the potential benefits of UQ carefully explained.
- It can be hard to define an acceptable or appropriate level of uncertainty in an analysis – should you choose 2σ or 3σ , for example? This choice is likely to be driven by external requirements, such as that of a safety case, but understanding what a level of uncertainty represents in practical terms and its consequence can often be difficult to establish.
- When establishing an overall safety position, it might be necessary to consider uncertainty from multiple analyses, each have their own separate, and potentially different assessment method (some may be BEPU, others conservative). In this case the overall uncertainty can be difficult to interpret and may lead to contradictory or an overly onerous or restrictive position (where too many uncertainties are concatenated on analysis and safety limits).
- Doing some sensitivity analysis and uncertainty quantification is better than doing none – perfection is neither economic nor possible for real problems. However, as long as a model is capable of reliably producing acceptable agreement with experimental data, making a judgement with some knowledge of the uncertainty, starting with a focus on the choice of uncertain inputs and defining their range/distribution, invariably adds to the understanding of a system’s behaviour.

4 Mathematical Methods for Propagating Uncertainty

4.1 Conceptual Statistical Framework – Frequentist vs Bayesian

Statistical methods generally fall into one of two categories: frequentist statistics or Bayesian statistics¹.

- In a **frequentist** approach the probability of a given event is based on a set of observed measurements of the event. In frequentist analysis, the parameters of a probability model (e.g. the mean and standard deviation of a normal distribution) have a fixed unknown value and are often represented by an estimated value and its confidence interval. Statistical (or null) hypothesis testing and p-values² are used to determine if a given event is statistically significant given the observed data (DeGroot and Schervish, 2012).
- In a **Bayesian** approach the parameters of a probability model are generally treated as uncertain quantities themselves, represented by probability distributions (Gelman *et al.*, 2014, Kruschke, 2015). The parameter values are calculated by combining assumed prior knowledge about the event, which may include subjective beliefs, with available data. During the calculation, the 'prior' distribution (or hypothesis) is updated, using numerical methods, with knowledge taken from additional data (or evidence) to define the 'posterior' probability distribution that an event will occur.

For example, suppose that an uncertain input variable follows a normal distribution, with the distribution parameters μ and σ representing the mean and standard deviation respectively. In a frequentist approach, the limited amount of measurement data collected defines a central estimate and confidence interval on μ and σ . As more data is collected the size of these confidence intervals reduce. In a Bayesian approach, μ and σ are treated as random variables defined by posterior probability distributions which are a result of the assumed prior distribution, updated for knowledge regarding the measurement data. As more data is collected, the analysis and posterior distribution is updated to reflect this.

Statistical software tools are available with implementations of many of the algorithms to perform frequentist and Bayesian statistical analysis, and either approach may be used depending on circumstance and the preference of the user. In many simple cases for characterising an uncertain input the results will be similar. Frequentist methods tend to be used more commonly, and Bayesian

¹ Choosing one or the other is loosely analogous to choosing whether to use the finite element or finite volume approach for discretising PDEs for CFD – the implementation details for any methods stem from this choice.

² p-values are an estimate of the probability that a particular result, or a result more extreme than the observed value could of occurred by chance. Crawley (2015)

Mathematical Methods for Propagating Uncertainty

methods can be more computationally expensive and can require more expert knowledge, but are becoming increasingly used. Bayesian methods can also be highly effective when there is little data available to inform decision making (for example to support expert elicitation).

In this technical volume the majority of the methods presented are frequentist in nature (specifically Monte-Carlo propagation of uncertainty). However, a description of how Bayesian approaches may be used to help characterise uncertain inputs is presented alongside a range of other methods in Section 4.2.4, and it is also possible to make use of some of the topics discussed within a Bayesian approach (Wu *et al.*, 2018).

4.2 Parametric Uncertainty Characterisation

Propagating uncertainty through a calculation methodology needs each identified source of input uncertainty to be appropriately characterised. In this process, each uncertain input is assigned a mathematical structure that describes the nature of its uncertainty, and specific numerical parameters for that structure defined (Roy and Oberkampf, 2011).

Aleatory uncertain inputs are most often characterised by a continuous PDF which represents the likelihood of the uncertain parameter as a function of different values. For example, the structure could be a normal distribution, and the parameters are the mean and standard deviation. However, more complex distribution forms could be used if required (for example, an asymmetric Generalised Extreme Value distribution (defined by location, scale and shape parameters) or the beta distribution (defined by two positive shape parameters in the interval $[0, 1]$).

Epistemic uncertain inputs should be conceptually separated from aleatory uncertain inputs. Epistemic uncertainties are commonly represented by an interval or take a limited number of discrete (or categorical) values, but some forms may also be able to be represented by a continuous PDF.

In practice, some uncertain inputs will be a combination of both aleatory and epistemic uncertainty and methods which account for both should be considered when characterising and propagating the uncertainty. For example, when modelling graphite, the porous structure of the material means that the thermal conductivity can only be represented by a probability distribution, and is hence aleatoric in nature. However, the thermal conductivity of graphite changes due to irradiation damage, and making measurements of irradiated graphite specimens is difficult, so only limited experimental data is likely to be available. Hence, there will be epistemic uncertainty on the parameters defining the probability distribution of the aleatory uncertainty for thermal conductivity at a given irradiation.

4.2.1 Considerations When Characterising an Uncertain Input

The method for characterising uncertain inputs should consider several aspects:

Mathematical Methods for Propagating Uncertainty

4.2.1.1 Where does the uncertainty originate?

The source of the input parameter uncertainty is significant in determining how it should be treated. For example, an uncertain input based on experiment can be defined by a well quantified PDF fitted to the available data, or it could be derived from relevant published experimental data. For example, in a model where a thermophysical property of a fluid is a key uncertain input, then a set of experimental measurements could be undertaken to define this value. The uncertainty on the experimental results can be used to characterise the simulation input uncertainty. The uncertain input may also be based on separate numerical calculations, modelling of a specific aspect of the system, formal expert elicitation methods, or as part of the PIRT process. A modelling choice or input may need to be based only on the user's (or an advisor's) expert judgement and there may be only limited evidence to support a credible quantitative uncertainty associated with the value.

4.2.1.2 Is the uncertainty aleatory or epistemic?

Uncertain inputs which can be described by continuous distributions are generally easier to propagate through to an output metric uncertainty. Aleatory uncertainties are usually specified by continuous distributions. It may also be possible to represent epistemic uncertainties by a continuous distribution and treat them in the same manner as aleatory uncertainties. Indeed, the distinction between aleatory and epistemic uncertainty may be dependent on the model. In this case the characterisation of the uncertainty as aleatory or epistemic should be a pragmatic choice, dependent on the purpose, to allow the uncertainty to be most easily represented by a continuous probability distribution. Kiureghian and Ditlevsen (2009) remark on the confusion that can occur when trying to assign an uncertainty to be aleatory or epistemic, or to determine when additional effort or knowledge can convert epistemic to aleatory uncertainty:

... we came to the conclusion that these concepts only make unambiguous sense if they are defined within the confines of a model of analysis. In one model an addressed uncertainty may be aleatory, in another model it may be epistemic. So, the characterization of uncertainty becomes a pragmatic choice dependent on the purpose of the application.

The distinction is useful for identifying sources of uncertainty that can be reduced in near-term, i.e. without waiting for major advances to occur in scientific knowledge. . . The distinction is also important from the viewpoint of transparency in decision-making, since it then becomes clear as to which reducible uncertainties have been left unreduced by our decisions. . . [and] can be further reduced, albeit at a cost that may not be justifiable.

More importantly, epistemic uncertainties may introduce dependence among random events, which may not be properly noted if the character of uncertainties is not correctly modelled.

Mathematical Methods for Propagating Uncertainty

4.2.1.3 What is the impact on the complexity and computational cost of the uncertainty propagation?

Epistemic uncertain inputs which are represented by an interval or a set of discrete choices can be more difficult to accommodate than aleatory uncertain inputs. For example, discrete distributions can lead to discontinuities in surrogate models which may make them hard to apply, requiring more costly full model evaluations instead. As a result, more complex analysis may be required to propagate the uncertainty through to a final output metric, as discussed in Section 4.8.

4.2.1.4 Is there combined aleatory and epistemic uncertainty?

When characterising an uncertain input both epistemic and aleatory uncertainty may need to be considered. Most often this is required if an uncertain input is random in nature, but there is epistemic uncertainty in how this random nature behaves. This might be because of:

Parameter uncertainty: The parameters required by the mathematical structure will themselves contain epistemic uncertainty, generally expressed by the mean and standard deviation of the coefficients. This represents the uncertainty on the average behaviour of mathematical structure in describing the behaviour of the uncertain input. The uncertainty in these parameters can be characterised and propagated through by UQ. However, the epistemic uncertainty on the parameters may be neglected if this is significantly smaller than the aleatory uncertainty and if the aleatory uncertainty is important for the output metric.

Functional form of the mathematical structure: If different functional forms can be used to define the mathematical structure of the uncertain input (including the aleatory uncertainty) there may be epistemic uncertainty to consider in the UQ. The best performing functional form should be evaluated where possible³, but alternatives considered as part of the UQ. Uncertainty in the mathematical structure would need to be treated as an epistemic uncertainty. This is particularly important if model extrapolation is required. Finally, it may be necessary to include epistemic uncertainty on the functional form and on the parameters discussed above in the UQ.

Extrapolation: If the uncertain input in the model is outside the range of the data used to determine its parameters, then extrapolation of the mathematical structure is required. This should be avoided if possible, because uncertainty can increase rapidly outside of the range of available data. It is likely that additional justifications that the mathematical structure chosen extends to the required domain will be necessary (for example by embodying the appropriate physics) and that the additional uncertainties that are created by this are considered.

Model calibration: If a model is 'tuned' to measurement data from the system being modelled via calibration to derive or adjust a parameter, this may make the uncertainty analysis more complex. Changes to the model inputs will inevitably lead to a change in the model calibration parameter. As a result, the parameter will likely need to be recalibrated against the measurement data for each model sample run.

In cases where there are combined aleatory and epistemic uncertainties, the uncertain input may be characterised by a p-box (described in Section 4.2.4.12).

³ For example, by examining the Akaike Information Criterion (AIC) (Vose, 2008) which is a relative measure of 'goodness of fit' for statistical models using the same data (with a penalty for increasing the number of coefficients).

Mathematical Methods for Propagating Uncertainty

4.2.2 Example of Considerations for an Uncertain Input

An example of a continuous uncertain variable is a thermophysical property of a material (for example, a thermal conductivity). In a NTH model, the mean value from a set of published experimental measurements may be used. However, the experimental uncertainty means that the true thermal conductivity is within a continuous uncertain range which must be parametrised. Two potential scenarios which could be considered are:

Aleatory uncertainty is not significant: This occurs if the material is homogenous and its manufacturing process is well controlled (such that a single value of true thermal conductivity value is valid for all the material) or the output metric of importance is only sensitive to the mean value. In this case the uncertainty on the mean thermal conductivity is only due to there being a limited amount of data to characterise it. Although, this may strictly be an epistemic uncertainty (it could be reduced with more measurement data) it is unlikely that further or improved experimental measurements are possible in the short term and the uncertainty on the mean thermal conductivity can be characterised by a continuous PDF. In the most simple case this might be via a normal distribution with central estimate given by the mean of the experimental data and a standard deviation characterised by the standard error on the mean.

Aleatory uncertainty is significant: This might occur if the material is non-homogenous or there is intrinsic variability in the component manufacturing (for example, the thermal conductivity of graphite in a large number of moderator bricks), and the output metric of interest is sensitive to extreme values of the aleatory uncertainty. In this case the uncertainty in the aleatory distribution must be characterised (i.e. how uncertain is the standard deviation of the distribution of thermal conductivity?) This is significantly more complex and requires both epistemic (uncertainties on the parameter estimates) and aleatory (the distribution width) uncertainties to be propagated to the output metric.

This is illustrated in Figure 4.1. In Figure 4.1a an input with aleatory uncertainty is characterised by a normal PDF fitted to the measurement data (the purple line). However, due to the limited amount of measurement data (represented by the histogram), there is uncertainty in the normal distribution that best describes the data, resulting in many possible normal distributions (the grey lines) which could also describe the underlying aleatory distribution. This represents epistemic uncertainty on the parameters of the normal PDF. If the output metric uncertainty is only dependent on the mean value of the uncertain input, then the confidence interval on the mean of the input can be used to characterise the uncertain input. However, if the entire aleatory uncertainty distribution is important for output metric uncertainty, then consideration also needs to be made of the epistemic uncertainty in the fitted distribution, and propagated through the UQ method. This uncertainty will likely be significant larger in the distribution tails compared to the central estimate. Figure 4.1b illustrates how taking more experimental measurements will improve knowledge of the aleatory distribution form and reduce the epistemic uncertainty on the normal PDF distribution parameters.

Mathematical Methods for Propagating Uncertainty

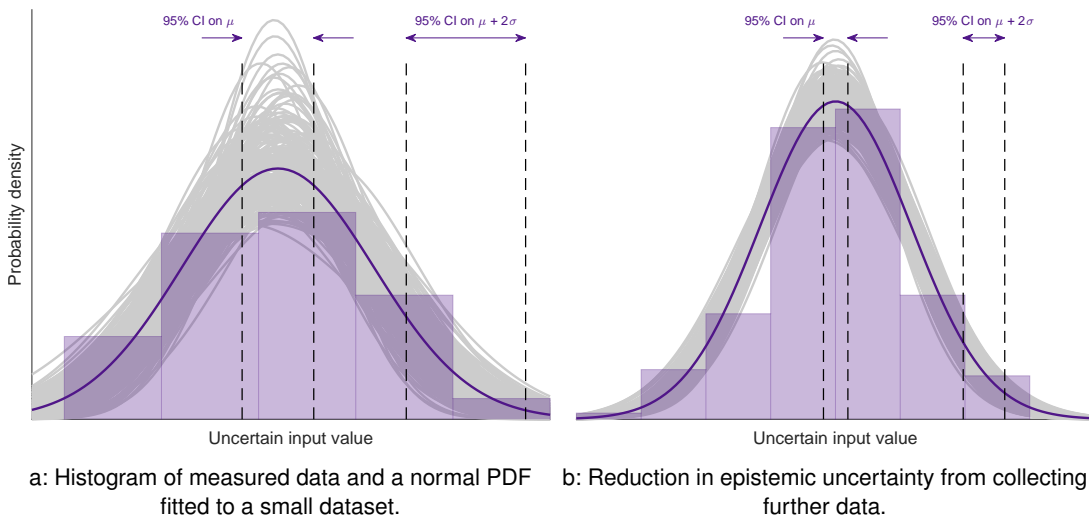


Figure 4.1: Example of an aleatory uncertain input and the impact that limited input data has on the uncertainty of defining the mean, μ , and standard deviation, σ , of a PDF of the input. The grey lines show a set of alternative normal PDFs that could represent the data, which result in epistemic uncertainty in the mean and 2σ of the distribution.

4.2.3 Central Limit Theorem

The Central Limit Theorem (CLT) is the most important result in statistics (Navidi, 2020) and states that the mean, \bar{x} for a large number (N) of independent samples from an underlying distribution, $f(x)$, will be normally distributed (Vose, 2008):

$$\bar{x} = \text{normal}(\mu, \sigma / \sqrt{N})$$

where μ and σ are the mean and standard deviation of distribution $f(x)$. This holds regardless of the functional form of the underlying distribution $f(x)$. The CLT also holds when calculating the mean from a large number of samples from different distributions, provided no one distribution dominates.

The CLT has two important consequences for an UQ assessment. Firstly, when characterising the uncertain inputs, uncertainty on the mean value can, in most circumstances, be described by a normal distribution. Secondly, the output metric uncertainty is likely to be approximately normally distributed, unless the underlying functional response of output metric in the NTH analysis is highly non-linear.

4.2.4 Characterising Uncertain Inputs

In general, an uncertain input can be described in a probabilistic manner (predominately aleatory uncertainty), by an interval (epistemic uncertainty), or by an imprecise distribution (combined aleatory and epistemic uncertainty, Roy and Oberkampf, 2011). Several common statistical techniques and methods to help characterise an uncertain input are discussed below.

Mathematical Methods for Propagating Uncertainty

4.2.4.1 Descriptive Statistics

The simplest method to characterise an uncertain input is to use descriptive statistical estimators. Supposing there are i independent samples of an uncertain input available (from experiment or other analytical models), and assuming that the data is normally distributed, then the mean and variance can be used to characterise the uncertain input. However, because there are only a limited number of data samples there is uncertainty in the estimators, commonly known as the standard error. These are equivalent to statistical fluctuations (or the standard deviation of) the estimator. Table 4.1 shows expressions for how to calculate the most common estimators and their standard error. In practice, it may be more appropriate and accurate to use numerical techniques (such as bootstrapping, discussed in Section 4.2.4.9) to find the uncertainty on statistical estimators, especially for the variance.

	Statistic	Standard error
Sample mean	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$	$\sigma_{\bar{X}} = \frac{S}{\sqrt{n}}$
Sample variance	$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ $= \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)$	$\sigma_{S^2} \simeq S^2 \sqrt{\frac{2}{(n-1)}}$

Table 4.1: Simple descriptive statistical estimators of a distribution of data and the standard error (or uncertainty) on the value as a result of having a limited sample of data (Ahn and Fessler, 2003). The estimate of the standard error of the mean, $\sigma_{\bar{X}}$, is applicable to data representing any probability distribution. The estimate of the standard error in the variance, σ_{S^2} , is applicable to data with a normal distribution – estimates can be made for other distributions, but require the kurtosis to be estimated (Cho *et al.*, 2005). Note that it is often preferable to calculate sample variance using the second variant, because it can be achieved in practice by accumulating running sums of X_i and X_i^2 for a value. Using the first variant requires all X_i values to be available at once, and so in a long transient calculation, for example, it would require data from all time points to be retained, because \bar{X} is only available at the end.

4.2.4.2 Analytical Uncertainty Propagation

Some uncertain inputs to a model may be calculated based on other measurements. For simple expressions it is possible to analytically characterise the uncertain input derived from the other measurements using the standard equation for error propagation. For a function $z = f(a, b, \dots)$ the standard deviation on the output z is given by using the partial derivatives expressing the sensitivity of the uncertain output to the input measurements

$$\sigma_z = \sqrt{\left(\frac{\partial z}{\partial a}\right)^2 \sigma_a^2 + \left(\frac{\partial z}{\partial b}\right)^2 \sigma_b^2 + \dots + \frac{\partial z}{\partial a} \frac{\partial z}{\partial b} \sigma_{ab}^2 + \dots}$$

where σ_i is the standard deviation of measurement i and σ_{ij} is the covariance between measurements i and j . Where the measurements can be assumed to be not covariant, then the ‘cross-terms’ are zero.

Mathematical Methods for Propagating Uncertainty

4.2.4.3 Regression Methods

Regression analysis can be used to find a functional relationship (f) between a dependent variable (Y_i) and one or more independent variables (X_i):

$$Y_i = f(X_i, \beta) + \epsilon_i$$

where β_i are the unknown coefficients in the function and ϵ_i is an error term for un-modelled features and random variable noise. The most common forms of regression modelling are:

- Ordinary (or linear) Least Squares Regression (OLS): A model where all the terms are a constant or a parameter multiplied by one or more independent variables raised to a power. Variable transforms can be performed to linearise certain functional forms.
- Non-linear Least Squares Regression: A model where the terms contain non-linear combinations of the independent variables. For example, exponential and trigonometrical functions.

In regression modelling the optimal set of coefficient values are found to match the data, normally by finding the global minimum of the sum of the square errors (making it closely related to 'curve fitting').

In NTH, regression modelling is commonly used to determine correlations between dimensionless quantities or dependency of physical properties on environmental conditions. For example, regression modelling could be used to determine a functional form between a thermophysical property (the dependent variable) and temperature (the independent variable) of a material based on a set of experimental measurements. There is extensive literature on performing regression modelling (DeGroot and Schervish, 2012) and it is included in statistical software packages. However, care should be taken that the underlying regression modelling assumptions hold⁴ to ensure the model accurately represents the observed data and that the model is not over-parametrised⁵.

The uncertainty in regression modelling is primarily characterised by the epistemic uncertainty in the fitted functional form, the epistemic uncertainty in the fitted coefficients (defined by the standard error in them), and the aleatory uncertainty in the observations. This is illustrated in Figure 4.2. The data is created from the (cubic) true function form, with the addition of Gaussian random noise (representing aleatory uncertainty). The dashed lines represent OLS fits using quadratic and cubic functional forms. The difference between the red and green lines represents epistemic uncertainty in the functional form. Neither fit is a particularly good representation of the true function around a value of 1 for the independent variable, because the aleatory uncertainty in the data (grey shaded

⁴ The key assumptions of regression modelling are: 1) The residuals (differences between each datapoint and the regression mode/curve fit) should be normally distributed, meaning that large deviations have a lower probability of occurrence 2) each data point is independent and an equivalent/comparable measurement of the same quantity and 3) the magnitude of the residuals does not vary as a function of the independent variables (homoscedasticity). In practice, regression models and methods for fitting them are generally robust to moderate deviations from these assumptions, and more complex methods exist to account for and correct deviations (such as generalised and robust least-squares).

⁵ An over-parameterised model can be one in which there are too many unknown coefficients for the amount of available data, and it is mathematically overdetermined. It can also be one where a regression function with more coefficients than are needed is used (even where a large number of data points are available). For example, fitting a high-order polynomial function to data that is inherently quadratic introduces coefficients that have no purpose. In both cases, an over-parameterised regression model will attempt to describe the noise in the data as well as the underlying physical trend, which can impair its performance. Methods are available to assess whether each coefficient is significant, and to add them progressively (stepwise regression) so that unnecessary terms can be removed or avoided.

Mathematical Methods for Propagating Uncertainty

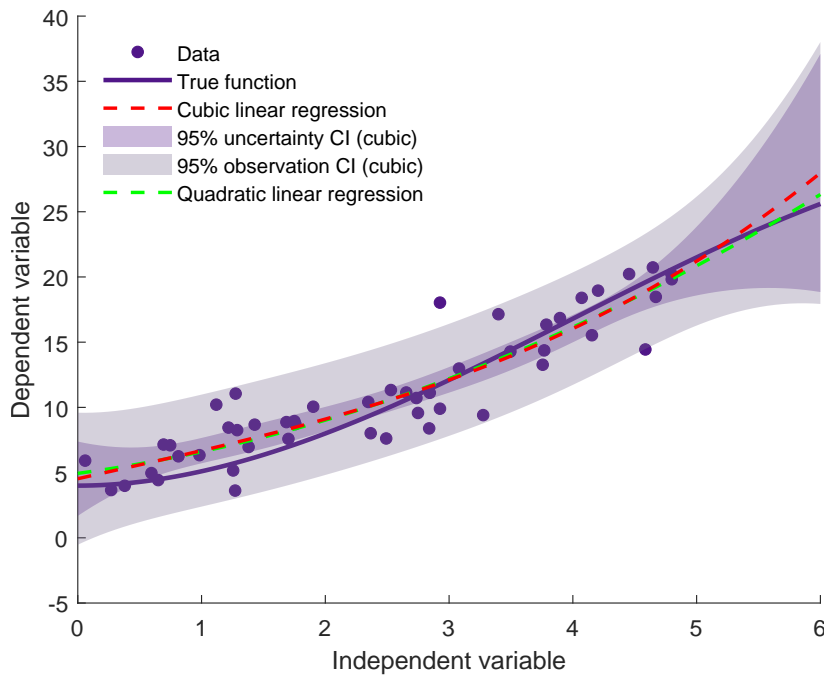


Figure 4.2: An example of the different forms of uncertainty in a linear regression model.

region) is (randomly) biased above the true function. The epistemic uncertainty in the cubic fit model coefficients is represented by the purple shaded region. A better fit (reduced epistemic uncertainty in model coefficients and closer agreement to the true function) would be achieved by having more input data points. There is a significant increase in epistemic uncertainty in the model fit predictions above a value of 5, where there is no data, highlighting the difficulties and risks in extrapolating regression models beyond the range of their fitting data. The necessity to include one or more of these uncertainties (possibly in combination) must be decided based on the considerations outlined above.

4.2.4.4 Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is a method for choosing the set of parameters for an underlying statistical model by calculating the set of coefficients which give the most likely chance the observed data occurs. It is based on evaluating the likelihood function for x_i data for a distribution function f and parameters θ :

$$L(\theta) = \prod_{i=1}^N f(x_i, \theta)$$

This likelihood function represents the probability of observing all of the data with parameters θ (DeGroot and Schervish, 2012). The maximum likelihood estimator (i.e. the set of coefficients $\hat{\theta}$ most likely to describe the underlying data for f) can be found by attempting to maximise the likelihood function. If it can be assumed that the data are independent samples from the same distribution, it is often more useful to work with the log-likelihood function for numerical reasons.

The benefit of MLE is that it can provide a consistent approach in parameter estimation across a number of different distributions, and comparing the AIC between different models can give a good indication of the best distribution form. It should be noted that at small sample sizes MLE can lead

Mathematical Methods for Propagating Uncertainty

to bias in the estimates.

As for regression modelling, there is extensive literature on performing MLE and it is included in statistical software packages. Furthermore:

- The parameter uncertainties found from MLE must be considered and propagated through the UQ if necessary.
- The form of the assumed distribution, f , also introduces uncertainty, particularly if the output metric is sensitive to the tails of the distribution which is poorly defined by the available data.

4.2.4.5 Kernel Density Estimation

Kernel Density Estimation (KDE) is a non-parametric representation⁶ of the PDF of a distribution of data. It is useful if the distribution is not easy to model via a parametric distribution (e.g. if it is bimodal) or if it is desirable to not make specific assumptions about the form of the data. For x_i data, the PDF of a KDE is calculated by

$$\rho(x) = \sum_{i=1}^N K(x - x_i, h)$$

where $K(x, h)$ is the Kernel, which is a positive function with a bandwidth h . Common forms of K are Gaussian, exponential and tophat. The results of a KDE can be sensitive to the functional form of K and, more importantly, the bandwidth. The bandwidth controls the level of smoothing applied by the Kernel function. Higher bandwidths smooth the data more, but can lead to important features in the data being removed. However, many analytical software packages (such as those described in Section 2.2.2) have automated methods for choosing appropriate defaults. Once the KDE has been calculated it can be evaluated at any value of x to propagate through the uncertainty analysis. Uncertainty in the KDE could be calculated via bootstrap methods.

4.2.4.6 Empirical Cumulative Distribution Function

A CDF can be directly calculated from a set of independent measurements of a quantity, this is known as the empirical CDF (eCDF). Uncertainty on the eCDF can also be found using methods such as Greenwood's formula and Kolmogorov-Smirnov bounds. Such methods are commonly used in survival/failure analysis (Lee and Wang, 2003) and can be used when there is a significant amount of available data or to better understand or analyse data prior to fitting a statistical distribution.

4.2.4.7 Bayesian Inference

The Bayesian inference process for characterising uncertain inputs is defined by two key-concepts (Kruschke, 2015, Gelman *et al.*, 2014):

1. Inference is the process of apportioning credibility (or likelihood) across a number of possibilities.

⁶ A non-parametric method is one which is not restricted to a defined distribution form. Instead, the method is numerical description based on the underlying data.

Mathematical Methods for Propagating Uncertainty

2. The possibilities to which the credibility is assigned can be described by parameters in a mathematical model.

The process for Bayesian inference is:

- Define a probability model of the system. This is a combined (joint) probability model including uncertainties in measured quantities, uncertainties in model predicted output metrics, and in the parameters defining the uncertain input. This is represented by appropriate prior distributions – user-chosen mathematical models with initial assessments of their credibility. This aspect of the Bayesian modelling process is often the hardest. It requires subjective prior distributions to be assigned with a subjective credibility which the outcome of the Bayesian inference may be dependent on.
- Inference (or updating) of the model to account for the known observations (the available data being used to construct the uncertain input). This involves numerical calculations (using Bayes' rule) to refine or update the credibilities to make them most consistent with the observed data.

The output of the Bayesian inference is a set of posterior probability distributions which describe the likelihood of occurrence of each possibility. These probability distributions can then be directly used to characterise an uncertain input for a UQ propagation. The outcome of the Bayesian inference should be assessed to check that they are consistent with the observed data; if they are not, then a different mathematical model may be required.

4.2.4.8 Expert Elicitation

Expert elicitation methods can be used where there is a lack of alternative data to make a characterisation of an uncertain input. For example, where experimental measurements are not possible or are too expensive or time consuming that they are impractical. The required output for the expert elicitation process can be the same as the other methods discussed, a mathematical characterisation of the uncertain input (Morgan, 2014).

The process for expert elicitation is:

- Define the uncertain input to characterise. The nature of the input to characterise will influence the expert elicitation process. For some inputs there may be a very limited number of experts to consult. For others, the spectrum of available experts may be much broader.
- Select appropriate experts with knowledge of the subject. It should be noted that the number of experts for a particular field may be limited or be close collaborators meaning the independence of different experts should be considered.
- Question the experts. These questions should be framed with the purpose of finding subjective probabilistic assessments of the uncertain input. This may be difficult, especially in circumstances where experts are unfamiliar or uncomfortable with making subjective probability assessments. The use of qualitative words should be avoided where possible, because their interpretation may differ substantially between experts and contexts⁷.

⁷ In a number of contexts that requires consensus on meanings and consistent application of terms, such as intelligence and climate science (IPCC, 2010), the concept of 'Words of estimative probability' is used, where specific terms indicate a band of numerical probability.

Mathematical Methods for Propagating Uncertainty

- Formulate probability distribution. The answers to the questions are used to form a subjective probability distribution of the uncertain input for the individual experts. Iteration with the previous step is possible, and initial responses of the experts may need further refinements to elicit their best subjective judgement. It is likely that the output probability distributions will be generated based on the experts' indirect knowledge and point estimates, rather than a direct knowledge of the distribution form and parameters. This can be captured by using three-point approximations (Keefer and Bodily, 1983) or bracket median methods (Smith, 1993) for example. A number of tools (for example Morris *et al.*, 2014) are available to support this activity.
- Combine results. The results of the experts are combined to form a final subjective probability distribution which defines the uncertain input.

Expert elicitation is particularly prone to contextual bias, where the experts are unintentionally influenced by background information. Furthermore, there is evidence that expert judgements tend to be over-confident (i.e. the outcome is probability distributions which are too narrow, see Hemming *et al.*, 2018). As a result, expert elicitation should generally be performed within a well defined framework or protocol to try and mitigate such bias. Example of such protocols include Delphi (Brown, 1967), Cooke (Cooke, 1991), IDEA (Hemming *et al.*, 2018) and SHELF (Gosling, 2018). One of the main differences between protocols is the level of interaction between experts. For example, in the Cooke protocol the subjective probability distributions of different experts are mathematically aggregated with no interaction between them. Alternatively, some protocols (such as SHELF and IDEA) employ behavioural aggregation where the output probability distributions are a result of independent assessment and facilitated discussions between experts. Mathematical aggregation protocols may achieve faster results, as there is no need for a consensus to be reached and facilitated discussions are not required, but often require some form of subjective ranking to be applied (via a mathematical formula) to produce the combined distribution.

The rigour and depth to which expert elicitation be adopted is dependent on the uncertain input in question and the resources available. Fully completing the protocols outlined above may be highly onerous, and it may be appropriate to adopt a more pragmatic approach to the expert elicitation process. For example, in its simplest guise, expert elicitation may involve consulting a single expert to set upper and lower bounds on an epistemic uncertainty which does not have dominant impact on the output metric.

Expert elicitation can also be used in combination with Bayesian updating methods, where it can form the basis of the assumed prior distribution to the analysis.

4.2.4.9 Bootstrapping and Jackknifing

Bootstrapping and jackknifing (Efron and Gong, 1983, Efron and Tibshirani, 1993) are numerical methods to estimate the uncertainty for fitted parameters and distribution coefficients by resampling from a limited underlying dataset and recalculating the required metric. Supposing that there is a dataset $X = \{x_1, x_2, \dots, x_n\}$, where each x_i is an set of independent observations of the same quantities from the underlying distributions, with a true quantity of interest F .

In bootstrapping a large number, N , of new resampled datasets are created $\{X_1^*, X_2^*, \dots, X_N^*\}$. Each resampled dataset is the same size as the original, so it will contain repeat values of the

Mathematical Methods for Propagating Uncertainty

same x_i and some x_i will be missing, but each is considered to be representative of what might have been observed if a different set of observations had been taken. For each of the N resampled bootstrap datasets the quantity of interest is calculated (such as the mean variance, or a parameter estimation via regression modelling). The result is a distribution of the quantity of interest $\{F_1^*, F_2^*, \dots, F_N^*\}$, which can be used to define the uncertainty on the input. The bootstrapping technique is powerful because it can be applied easily to both simple and complex quantities of interest. Because bootstrapping is limited to the set of input measurement data, it cannot represent uncertainty due to systematic observation uncertainties, or if there is a subset of important outliers which have not been observed.

In jackknifing, the quantity of interest is estimated n times, but excluding the x_i data point from each subsample calculation. This is a systematic resampling, compared to the random resampling of bootstrapping. Where the bootstrap gives an indication of the uncertainty on the quantity of interest, jackknifing can give an estimate of the bias (i.e. whether the dataset is likely to over or under-estimate the parameter). If the jackknifing shows a likely bias in the sample statistic, it may be necessary to take this into account when characterising the uncertain input.

4.2.4.10 Interval Range

For purely epistemic uncertainties the uncertainty can be characterised by an interval range (Ferson *et al.*, 2007). Interval analysis defines an upper and lower bound associated with a quantity but little else is known about a quantity. Mathematically, a quantity X that can take any set of values within the set interval value is defined as

$$X = \{x : a \leq x \leq b\} = [a, b]$$

The interval may be based on quantitative information that defines the bound (for example using the statistical techniques outlined in this section to derive the uncertainty on the mean of a normal distribution) or on a subjective assessment of the bounds (e.g. from an expert elicitation process). The generation of further information (e.g. more measurements) could be used to reduce the size of the interval. Depending on the importance of the input uncertainty and amount of information available in defining the interval, it may be appropriate to apply a discrete (and justified) safety factor to the interval to ensure it covers the underlying uncertainty⁸.

An extension of interval range analysis is the representation of a uncertain input by fuzzy logic or numbers in which the the level of possibility of a value is constructed via a set of nested intervals to produce a fuzzy set (Ross, 2010). The fuzzy number of a parameter a given value may take any real value between 0 and 1 inclusive and represents the possibility that the uncertain input lies within a fuzzy set (or the degree of uncertainty that the value belongs within the set).

⁸ Rigorous algorithms exist which take interval values inputs and provide bounds on the accumulated rounding errors, approximation errors, and propagated uncertainties (Moore *et al.*, 2009), but need to be embedded throughout the calculation route, and are not covered in more detail here.

Mathematical Methods for Propagating Uncertainty

4.2.4.11 Discrete or Categorical Probabilities

As noted in Section 2.1.2, for some epistemic uncertainties the NTH model may be represented by a set of discrete uncertain possibilities, such as the choice of turbulence model. This is representative of uncertainty in the underlying model form and can be modelled by a set of discrete input options, each with an assigned probability. The overall probability of these options should sum to one. As such, these types of uncertainty can be modelled by a categorical Probability Mass Function (PMF).

The probabilities of the different options may be quantitative or subjective and it should be noted that the discrete options may not fully represent the whole uncertainty in the model formulation. However, characterisation of key modelling choices via a set of discrete options may allow important aspects of model formulation to be propagated through the UQ and incorporated in the reported uncertainty on the output metric.

4.2.4.12 Probability Bounds Analysis

The characterisation of an input may need to be the combination of aleatory uncertainty (treated by probabilistic methods) and epistemic uncertainties (represented by discrete values or intervals). As discussed above, the parameters describing a probability distribution of an uncertain input are themselves uncertain (due to a limited amount of input data), and this needs to be considered in the overall UQ (Roy and Oberkampf, 2011).

A combination of probability theory and interval arithmetic can give rise to probability boxes (p-boxes, Tucker and Ferson, 2006), which is similar to a CDF but have a finite width representing the epistemic uncertainty (this is also known as imprecise probability). A p-box comprises of two non-intersecting CDFs which represent the upper and lower bound CDFs which contain the family of all possible probability distributions. This is illustrated in Figure 4.3 which shows how a set of CDFs can be combined to give an overall uncertainty. A single CDF is the realisation of one possible aleatory uncertainty distribution and the set of CDFs incorporates the impact of interval or discrete epistemic uncertainties. The width of the resulting p-box (Figure 4.4) represents the range of parameter values that are possible for a given cumulative probability level, whereas the height of the p-box represents the range of interval-valued cumulative probabilities associated with a given parameter value. In the limit that there is little epistemic uncertainty then the p-box degenerates to a precise CDF representing just the aleatory uncertainty

This is known as Probability Bounds Analysis (PBA) or Interval-Valued Probability (IVP), and it allows for a comprehensive description of both forms of uncertainty for an uncertain input. It is useful whenever the uncertainty about the distributions can be characterized by interval bounds about their cumulative distribution functions (Tucker and Ferson, 2006). The p-boxes represent the uncertainty about imperfectly known input distributions and these can be propagated through the model to identify a bounding p-box which encapsulates the entire uncertainty on the output metric. This approach has been applied to structural analysis (Choudhary *et al.*, 2016) and is the suggested starting point for combined uncertainties.

For example a p-box may be used to represent an uncertain input which is defined by a distribution fitted to measured data via MLE. The MLE can return an uncertainty range for the mean and standard deviation of the input and using this information the p-box can be calculated which envelopes

Mathematical Methods for Propagating Uncertainty

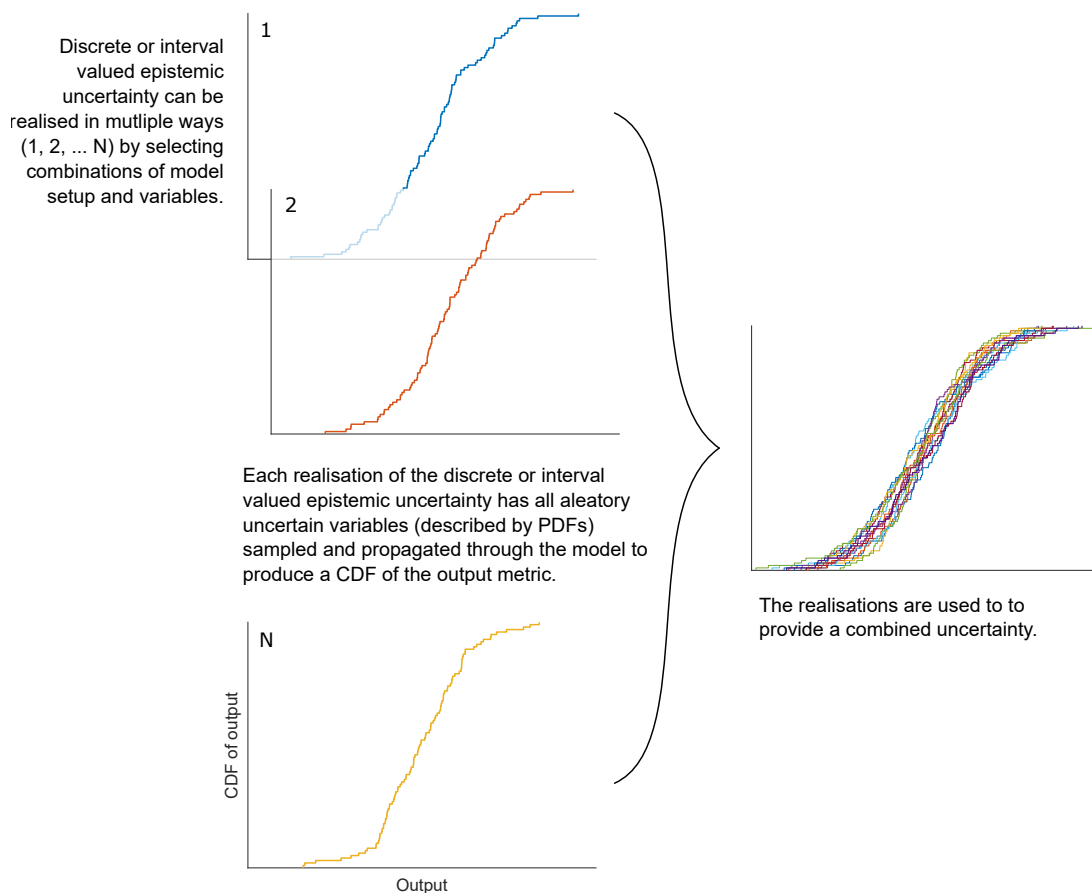


Figure 4.3: Example schematic showing propagation of probabilistic and discrete/interval uncertainties through a model. Each CDF on the left represents the uncertainty on an output metric for a given realisation of the epistemic/discrete uncertain inputs. These different realisations are combined to give the overall uncertainty on the output.

all the normal distributions within these intervals. Likewise, if the distribution form is uncertain a p-box can be estimated from simple statistics of the data sample such as the mean, minimum value and maximum value. A p-box may also be calculated from an empirical CDF, for example using Kolmogorov-Smirnov bounds. Further details on how to compute p-boxes can be found in Tucker and Ferson (2003).

4.2.4.13 Dempster-Shafer Theory

Dempster-Shafer Theory (DST) is an alternative to the PBA method for analysing epistemic uncertainties. DST is a mathematical theory of evidence which allows one to base degrees of belief for one question on probabilities for a related question (Shafer, 1976).

In DST a belief function is formulated and associated with a set of possible values which define the degree of belief that the value set will occur⁹. The belief in a value set may be a combination from multiple independent sources of evidence, formulated using a combination rule. A belief function is defined for all possible sets of values and the input uncertainties, as defined by the DST belief

⁹ The degree of belief in a set is different from a traditional probabilistic assignment. For example, a belief in an outcome of zero does not imply it is impossible, as it would for a traditional probability, only that there is no reason to believe it did occur.

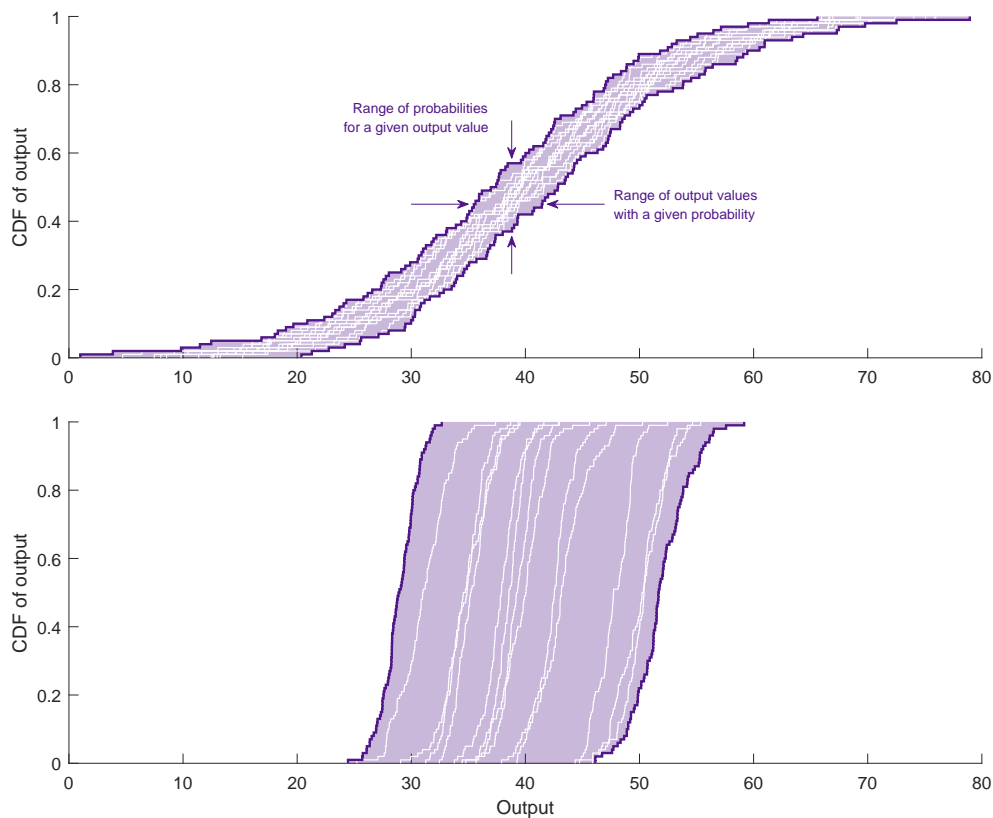


Figure 4.4: Example of two p-boxes using the same CDFs as Figure 2.3, where the 15 normal distributions are each sampled 100 times.

functions, can then be used to propagate epistemic uncertainties through an UQ calculation using direct calculation or surrogate models. DST is not typically used in the field of nuclear energy, but Darby (2018) demonstrates examples of it in the context of epistemic uncertainty in safety and security issues for nuclear weapons.

4.2.4.14 Correlation Between Inputs

It is possible that there is some correlation between uncertain inputs, and such correlations should be propagated through the UQ. For example, if a fluid composition is uncertain, this will affect its viscosity, thermal conductivity etc, so varying the composition and deriving the consequent common-cause variation in thermophysical properties (if this is known) is the most appropriate action, rather than allowing the properties to vary independently.

Ideally, these uncertain inputs into the NTH model would be expressed as joint PDFs which describe how sampling from one uncertain input influences others. A covariance matrix can be used to express the linear relationship between the variables. The correlation is calculated by normalising the covariance by dividing by the standard deviation of the variables. This results in a dimensionless quantity between -1 and 1 which can be used to compare variables which are expressed in different units.

When examining correlation in UQ it may also be necessary to consider:

- Correlations that exist between the uncertain input and a feature of the output metric. For

Mathematical Methods for Propagating Uncertainty

example, when measuring the thermal conductivity of a material as a function of temperature, the error in the input (the average temperature of the sample) may be correlated with the error in the temperature difference across the sample, that is used to estimate the output (conductivity).

- Whether it is the uncertainties on inputs which are correlated, rather than the values of the inputs themselves.

In practice, it is likely that there is insufficient information to adequately define the probability distributions for correlated inputs (Roy and Oberkampf, 2011) and it may be necessary to treat them as independent variables in a practical engineering implementation.

Furthermore, correlations between uncertain inputs can make propagation of them through the model more complex. For example, standard LHS in experimental design (discussed below) can be difficult when considering correlations between inputs, although specific methods have been developed to account for this (Sallaberry *et al.*, 2008). It is easiest to include correlations between uncertain inputs in simple Monte-Carlo methods of uncertainty propagation.

4.3 Sampling or Design of Computer Experiments

Computational experimental design provides the framework to statistically sample uncertain input parametrisations to capture the model response to them. Each statistical sample represents a single computer ‘experiment’ (simulation) to be performed. Effective experimental design optimises the number of model simulations that need to be performed (to keep the computational effort achievable) whilst ensuring that sufficient information is extracted to understand the effect of uncertainty in the input parameters on the overall model uncertainty.

An experimental design needs to consider:

- The dimensionality of the parameter space. For uncertainty or sensitivity analysis, this involves identifying the parameters that will be varied.
- The range and shape of parameter space to cover in each dimension (the interval or PDF for each parameter).
- Correlations between parameters.
- The total number of computer experiments which can be performed. This is linked to the cost of each simulation in terms of time and resources (simulation run-times, number of processors, memory and storage needed).

If the cost associated with individual evaluations of a model is so large that an insufficient number of calculations can be performed, then it indicates that using a surrogate model approach (Section 4.4) may be warranted, as long as the number of full-order simulations needed to construct the surrogate is relatively low. Experimental design can also be used to determine the computer model simulations required to adequately define the surrogate model. It is also the case that for costly models, sensitivity analysis cannot necessarily afford to be performed on a separate, specifically chosen set of model simulations, but must make use of the same set that are performed for the purposes of uncertainty quantification.

Mathematical Methods for Propagating Uncertainty

Much of the available experimental design literature is aimed at the design of physical experiments (as opposed to computer simulations). Whilst the two have much in common, computer simulations generally have the advantage that there is no notion of random error, i.e. if a computer simulation is run twice, it will (generally) produce the same result both times, unless there is a deliberate and controlled use of random numbers. Therefore, it is not necessary to perform repeat calculations with exactly the same input data.

Some of the experimental design techniques described in this section are most often applied to physical experiments (e.g. factorial design, central composite design or Box-Behnken design) whereas others are more commonly applied to computer simulations (e.g. random sampling design or stratified sampling design). In some contexts, the description Design and Analysis of Computer Experiments (DACE) is used to distinguish computer experiments from physical DoE (Adams *et al.*, 2020, Chapter 4).

Methods for performing experimental design can be broadly categorised as those based on selecting from a structured grid or array of locations in the sampled space, and those that have a less structured spatial relationship. In each case, where k denotes the number of parameters being sampled, and therefore the number of dimensions of the sampled space. NIST (2021, Section 5.3) provides details of these methods, both in the context of experimental design for physical experiments and computer simulations, and examples of the methods discussed are shown in Figures 4.5 and 4.6.

4.3.1 Structured Grid Based Methods

4.3.1.1 Factorial

In factorial design, each parameter (or 'factor') is allowed to take the value of a specific level. For example, in a two-level factorial design each parameter might take the maximum and minimum of its uncertainty interval. If all combinations of input parameters are accounted for ('full factorial design'), then the experiment would need to be run 2^k times. Setting up the experiment in this way generates an orthogonal array, eliminating correlation between the estimates of the main effects and interactions. Whilst conceptually simple, this method of sampling becomes prohibitively costly if the uncertainty in a large number of factors is being investigated.

One solution to the problem of the requirement for a large number of experimental runs is to perform a fractional factorial design, carefully choosing only a fraction of the runs specified by the full factorial design (for example 25 % or 50 % of the total runs). This is done by only choosing the main effects and ignoring some interactions between parameters. NIST (2021) and Andres (2002) provide in-depth discussion of factorial design methods.

All two-level fractional factorial designs are based on using only the maximum and minimum of each parameter's uncertainty interval, and not any values between them. This means that a surrogate model fit to these samples may not adequately capture the effects of variation within these intervals.

Mathematical Methods for Propagating Uncertainty

4.3.1.2 Central Composite

A Box-Wilson Central Composite Design (more commonly referred to as a central composite design) consists of the following parts:

1. A complete or fractional 2^k factorial design (the factorial portion).
2. A central point located at the median of each parameter uncertainty interval.
3. A set of 'star' points that are the same as the central point, except for one factor that is varied between values above and below the median (typically outside of the two factorial levels in the factorial portion), so they lie on the co-ordinate axes of the parameter space. There are twice as many star points as there are parameters in the design ($2k$).

Therefore, the total number of design points in a central composite design is $n = 2^k + 2k + 1$. These additional points (compared to the two level full factorial design) allow for the estimation of the second-order polynomial equation, whereas the two-level factorial design is limited to linear relationships between variables. More details are provided by NIST (2021).

A simpler, related method is a centred parameter study, where the factorial points are not used, and samples are only chosen along the axes of the parameter space (like the star points). Choosing more than three values along these axes allows for greater resolution of the non-linearity of the response, but this method only captures univariate sensitivities of the outputs to parameters (i.e. no combined variable effects).

4.3.1.3 Box-Behnken

In the Box-Behnken Design, experiments are run at each of the edges of the parameter space and at the centre of the space. Three levels are therefore required for each parameter.

The main advantage of the Box-Behnken design is that, compared to the full factorial design and the central composite design, fewer experimental runs can be required to capture the parameter space, depending on the number of parameters. However, unlike the full factorial and central composite designs, the sampled space is smaller and provides a poor quality prediction in the 'corners' of the design space (NIST, 2021).

4.3.1.4 Smolyak Sparse Grids

A related method for reducing the number of evaluations needed from a 'full tensor product' (all combinations of values of input variables, where each is sampled a specified number of times, s , giving s^k samples on a k dimensional grid) are sparse grids (or Smolyak sparse grids, Gerstner and Griebel, 1998). They are applied, for example, in Dakota (Dalbey *et al.*, 2020) to sample a structured reduced set of uncertain inputs from a complete k -dimensional grid for surrogate model UQ methods based on orthogonal polynomials (Section 4.4.2).

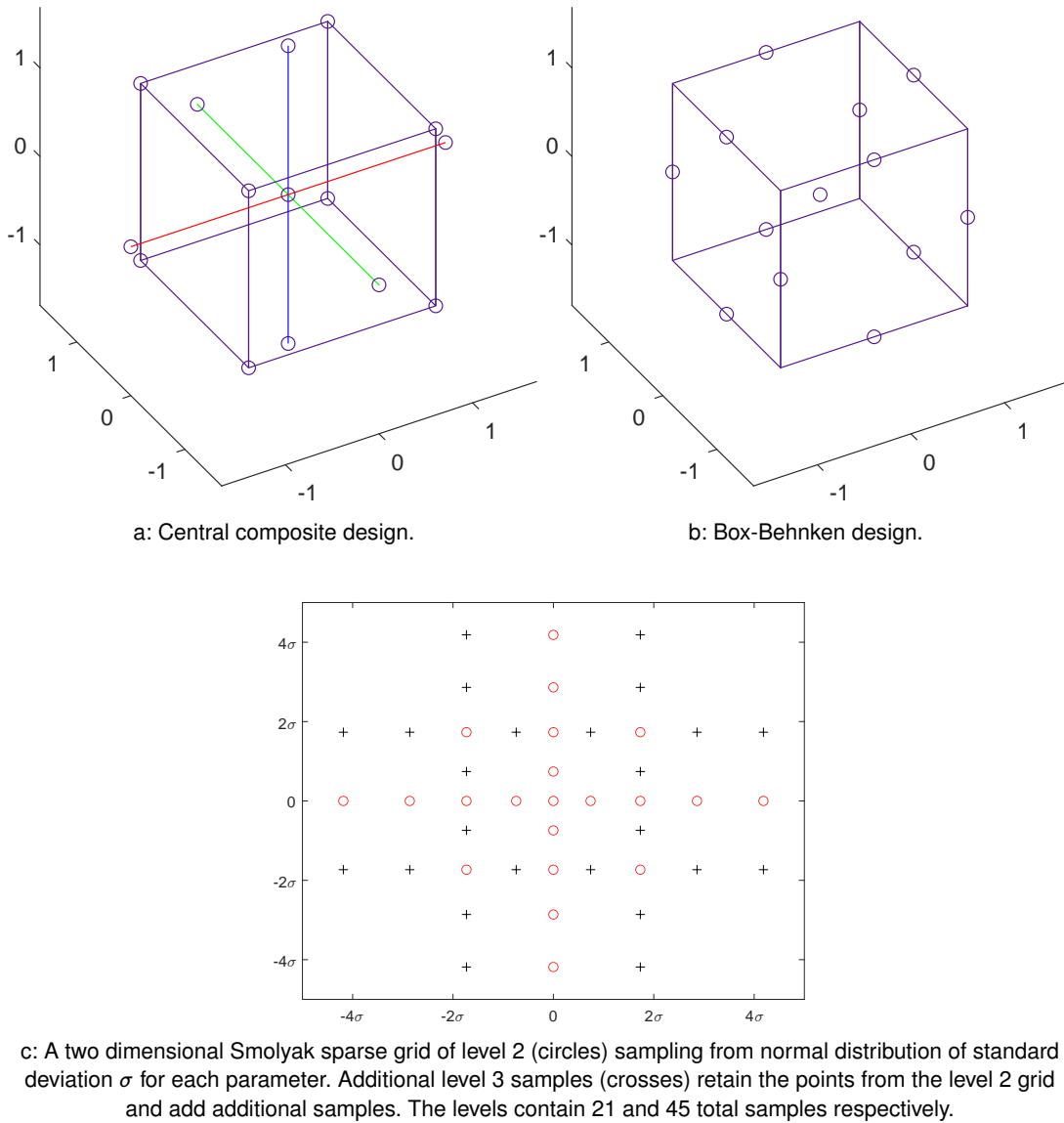


Figure 4.5: Structured sampling methods.

4.3.2 Unstructured Methods

4.3.2.1 Random Sampling

A random sampling experimental design is performed by drawing ‘Monte-Carlo’ samples from the parameterised distributions of the input variables. In a random sampling design, all of the parameter space has an equal probability of being sampled. The Monte-Carlo samples could be independent, but in more complex designs correlation between input variable sets can be accounted for using joint probability (or multivariate) distributions (Giunta *et al.*, 2003).

A random sampling experimental design is the simplest to envisage and perform, and (given sufficient sampling) results in an unbiased estimate of the mean response. However, it may require

Mathematical Methods for Propagating Uncertainty

a large number of samples to adequately cover the extremes of the input distributions. This is of particular concern if the uncertainty quantification is concerned with large confidence intervals (e.g. 99.9%) or if some input parameter distributions have long ‘tails’ which are significant (Andres, 2002).

A random sampling experimental design may be applied to an appropriate surrogate model, to use the faster surrogate to rapidly perform the large number of evaluations.

4.3.2.2 Stratified Sampling

Stratified sampling design aims to ensure a more uniform sampling over the design space compared to random sampling design. LHS is a form of stratified sampling design that is used extensively (and often the terms are used interchangeably in the literature). In LHS, if n samples are requested, then the range for each parameter is divided into n ‘bins’ of equal probability. Then, n samples are randomly selected such that exactly one sample is randomly placed inside each bin for each parameter.

The construction of the LHS design can give a more accurate characterisation of a response (i.e. narrower confidence intervals on quantities like the mean and standard deviation) than a simple random sampling design for an equivalent number of samples. However, it is also possible for conditions required by LHS to be satisfied whilst being poorly arranged in the sample space. For example, if all the LHS samples in Figure 4.6 were arranged along the diagonal, this would lead to poor coverage of the parameter space (Bhattacharyya, 2018).

Variations of stratified sampling are available, such as orthogonal array sampling (Giunta *et al.*, 2003), as are generalisations of LHS (Shields and Zhang, 2016, as applied, for example, by Banyay *et al.*, 2020).

4.3.2.3 Low Discrepancy Methods

Purely random sampling inevitably generates dense clusters of samples. Quasi-Monte Carlo methods produce deterministically generated sequences of samples, that have the appearance of being random, but are intended to cover the sampling space with more uniformity (lower discrepancy). Halton and Hammersley sequences are examples of these methods.

Another low discrepancy method is a centroidal Voronoi tessellation, where the sample points represent the centroids of surrounding space filling polygons. The positions of the samples are iteratively adjusted to achieve an approximately uniform distribution of polygon shapes and volumes. This is similar to the operation of some automated tetrahedral and polyhedral mesh generators¹⁰ used for CFD.

¹⁰ Often mesh generators construct 3 dimensional Delaunay triangulations, which are the ‘dual’ of the Voronoi tessellation. Methods used for sampling operate in k dimensions, giving the higher dimensional equivalents of tetrahedra and polyhedra.

Mathematical Methods for Propagating Uncertainty

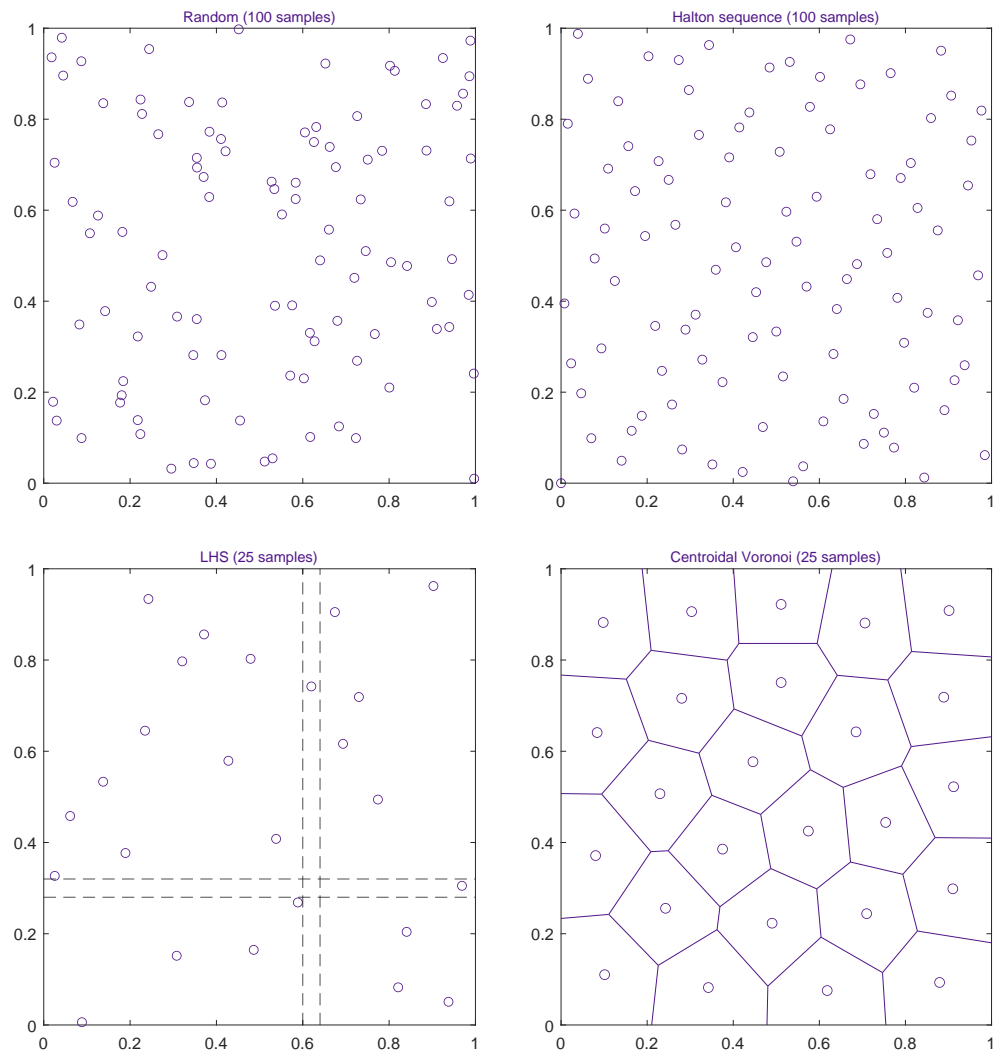


Figure 4.6: Unstructured sampling methods. The clustering of points characteristic of random sampling is seen, and contrasted with the more even spacing of the Halton sequence. The bin width for the LHS example is marked, showing that there is only one sample in each bin for each parameter dimension, although the coverage of the parameter space is not as uniform as achieved by the centroidal Voronoi method.

4.3.3 Selecting a Method

The appropriate sampling or DoE method to choose depends on the number of parameters being sampled, the range and shape of their variation, the cost and behaviour of the model that is being used to evaluate them, and the objective of the calculation:

- For an initial sensitivity assessment, a centred parameter study or central composite design may be useful to screen the relative effects of a larger number of parameters and establish the expected range of outputs. A centred parameter study with a larger number of samples for each parameter can identify finer structure and non-linearity in the results.
- Random number generators can easily draw samples in proportion to the probability density from any form of input distribution (they do not need to uniformly sample the parameters). However, because random or LHS sampling methods draw their samples from PDFs of their

Mathematical Methods for Propagating Uncertainty

uncertain parameter in a known way, the inputs are linked to the probability of the outputs in a way that permits statistical rigour in their interpretation, which may be necessary in some UQ applications.

- Low discrepancy methods provide more uniform coverage of the parameter space, but do not possess these more precise statistical properties. For samples being used to construct surrogate models, however, it is more important that the number of experiments performed and range of parameters used are sufficient to avoid significant distortion of the distribution of output metrics produced by the surrogate. For example, a response surface fitted to the samples from an experimental design with insufficient coverage of the parameter space may lead to an inappropriate functional form being chosen (e.g. a quadratic relationship may be fitted when in fact a higher order polynomial would have been more appropriate). This may introduce systematic uncertainty that is not accounted for in the calculated output metric uncertainty distribution. For a response surface application adequate samples are needed within the central region of parameter space, in which case, the Central Composite and Box-Behnken methods are likely to perform poorly.
- Building particular kinds of surrogate model may benefit from the structured arrangement and known resolution in each direction offered by sparse grid methods.

Guidance on which method to choose is difficult to formulate, other than to recommend that several options are assessed, that the coverage of parameter space is considered, and that the characteristics of the model response are evaluated carefully at each stage to choose where and how to concentrate attention.

4.4 Surrogate Models

The widespread availability of computing power has led to complex, high-fidelity numerical simulations (such as CFD) of physical phenomena being widely developed and applied. However, many situations during the design and licensing of a nuclear reactor motivate exploration and optimisation across a large parameter space. In this situation, high-fidelity computer simulations are time consuming and costly. For example, complex coupled multiphysics simulations (including structural or neutronic models for examples) may take many days to evaluate (potentially using hundreds or thousands of CPU cores) and involve solving equations for millions of degrees of freedom (grid/mesh points).

In some situations it is beneficial to develop simpler models which retain a sufficiently accurate representation of the important underlying physical phenomena. These simpler models are often referred to as surrogate models, response surface models, metamodels, fully equivalent operational models or emulators. Surrogate models have applications beyond SA and UQ; because they are fast to evaluate, they can also be used where performing computationally intensive simulations is impractical, such as design optimisation or real-time simulation. An overview of the process of using a complex model of reality to build a simplified surrogate model, and their typical applications, is shown in Figure 4.7.

Surrogate models can be broadly classed into three types:

- Simplified physical models: a lower fidelity physical approximation of the high-fidelity model.

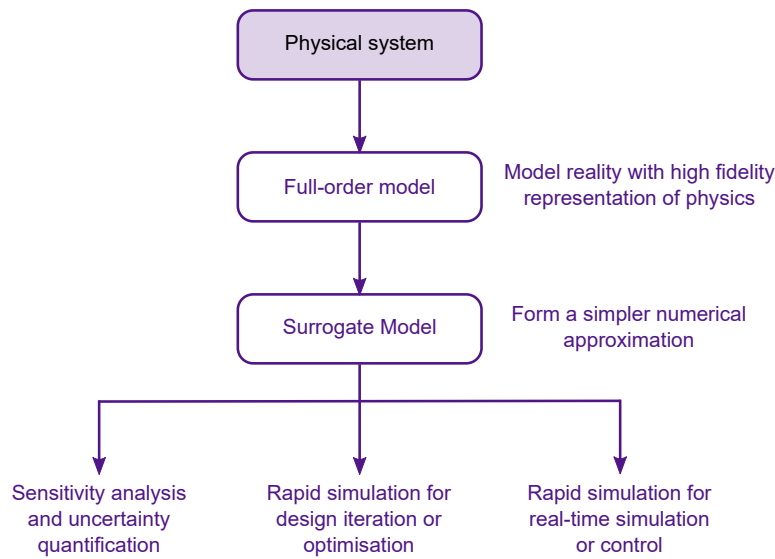


Figure 4.7: Overview of the use of surrogate models.

- Data-fit models: a response surface or emulator fitted to a sampled dataset from (ideally a small number of) high-fidelity model simulations.
- Reduced Order Models (ROMs): the high-fidelity model is projected onto a basis of reduced dimensionality compared to the original solution space.

Surrogate models use experimental design methods in two ways – it is necessary to choose the points in the input parameter space for the full-order model evaluations needed to build a surrogate model. Once the surrogate has been built and validated, it can be used to produce ‘emulated’ simulation results, generated using a different method of choosing samples, to provide the functionality of interest.

4.4.1 Requirements of a Surrogate Model

A surrogate model is, by definition, a simplification of the full NTH model of the physical system of interest. It cannot, therefore, capture all of the physics of the high-fidelity model. This can be expressed mathematically as

$$y = f(x) = f'(x) + \delta$$

where $f(x)$ is the high-fidelity thermal hydraulics model, $f'(x)$ is the surrogate model and δ is the approximation error associated with the surrogate model. Despite the modelling simplifications, a surrogate model can be a useful tool for several purposes, including for uncertainty quantification, if it meets certain requirements:

- The approximation error relative to the full-order model should be known (or bounded), and acceptably small in the parameter ranges of interest, and for the required purpose. For example, deriving accurate sensitivity information from a surrogate model may require there to be smaller errors than for other uses.
- The direction and approximate magnitude of the response of the surrogate model to changes

Mathematical Methods for Propagating Uncertainty

in the input variables must be consistent with underlying physics of the system¹¹.

- It must calculate the output metrics required for the model application.
- It must be sufficiently computationally efficient for the surrogate model application.

To demonstrate that the surrogate model fulfils these requirements, it is important to compare its output against the full-order model. For example, the surrogate model could be fitted to a 'training dataset'. The model performance should then be demonstrated on a different 'test dataset', not used in its construction, within the parameter space of interest to demonstrate whether approximation error is suitably small. It is also possible to use cross validation, which involves repeatedly fitting and testing the surrogate model against different subsets of the available dataset, to assess its accuracy and to check for bias.

The error in the surrogate model can be determined by examining the statistics output in the model fitting process (for example, R^2 , the RMS error, the mean absolute error or the maximum absolute error) and should be negligible compared to the uncertainties being propagated.

4.4.2 Types of Surrogate Model

4.4.2.1 Simplified Physical Model

One option for a surrogate model is to use a lower fidelity thermal hydraulics code that captures the important physics of the full-order model, whilst being less computationally expensive. For example, the full-order model could be a detailed CFD model, however a porous model or other coarse grid representation (see Volume 2, Sections 3.4.3.3 and 4.2) may offer substantial computational savings while retaining much of the behaviour of importance. Through the use of a 'test dataset', the simplified representation can be compared to the detailed CFD analysis to establish that it captures the relevant physical phenomena and allows the output parameter of interest to be calculated accurately with only a small approximation error.

A simplified physical model is likely to give the most accurate replication of the full-order model when extrapolating outside of the dataset used to derive or validate the surrogate model, because it is physically based. However, it is also likely that it will be more computationally expensive than some of the other techniques discussed below and may be more onerous to develop.

A simplified physical model may also be included as a part of a full-order model to reduce computational expense. For example, a porous representation may be included in regions of a CFD model to reduce computational cost for experimental calculations. A surrogate model, using one of the techniques outlined below, may then be built from this to perform more detailed uncertainty quantification.

¹¹ For example, in a system comprising an ideal gas, where $P/\rho = RT$, then if T increases, but the surrogate does not produce a physically compatible change in P or ρ , it is invalid. In this simple example, the expected behaviour is easy to predict and assess – for complex, realistic models and more subtle effects, it may not be.

Mathematical Methods for Propagating Uncertainty

4.4.2.2 Correlation-based Model

Correlations are the simplest form of a simplified physical model. They synthesise a particular idealised set of behaviour to a (usually) analytical expression linking non-dimensional numbers representing the flow or heat transfer regime to responses of interest (e.g. pressure drop and heat transfer). They are often based on experimental data and can have an underlying physical basis to their algebraic relationship (although this is not a requirement) and are likely to have coefficients derived via regression methods (Section 4.2.4).

Correlations are key to all aspects of system codes (a system code model for a complex two-phase flow accident could have hundreds of correlations embedded within it) and they are often used to provide boundary conditions for CFD models.

Volume 2, Section 2.1.2.3 discusses heat transfer correlations in more detail, and Volume 5, Section 3.1 provides guidance specific to liquid metals, but also raises a point about a source of uncertainty that may be easy to overlook. Correlations are often intended to create a relationship suitable for an 'ideal' representation of a particular geometrical configuration and flow (such as a tube bundle). However, any given experiment has systematic biases in operation. The variation seen in results for the same basic geometry highlights this, and correlations that combine all of the available data gain universality from doing so, but their specificity for reproducing an individual dataset is reduced (although this is hard to separate from the contributions made by measurement uncertainties).

4.4.2.3 Polynomial Response Surface Model

A response surface model maps input variables to the model outputs. Typically a full-order model is run to develop a dataset of predicted output metrics for a chosen set of input (or predictor/regressor) variables chosen by an appropriate experimental design method (Section 4.3); a response surface is then fitted to this data using standard regression techniques. The response surface is typically a first-order or second-order polynomial:

$$y = \beta_0 + \sum_{j=1}^k \beta_j x_j + \sum_{j=1}^k \sum_{i=1}^k \beta_{ij} x_i x_j + \delta,$$

where y represents the response, x_i represents the input variables, β_i are the regression coefficients, k is the number of input variables and δ is the error. Regression is used to determine the 'best-fit' coefficients, with the aim of minimising the response surface error, δ .

The response surface can then be interrogated for different input values to understand the impact on the required output metrics. Care should be taken near to the boundaries and particularly when extrapolating outside of the input range of the dataset to which the response surface has been fitted. It is likely that the response surface error will increase significantly outside of this dataset.

Mathematical Methods for Propagating Uncertainty

4.4.2.4 Polynomial Chaos Expansion Model

In a PCE surrogate model the output is expanded into a series of orthogonal polynomials, which are a function of the uncertain inputs x :

$$y \simeq f'(x) = \sum_{n=0}^P c_n \phi_n(x)$$

where ϕ_n are the (multivariate) polynomials of order, p , and c_n are the expansion coefficients. The number of terms in the expansion is $P + 1 = (k + p)! / k! p!$, where k is the number of uncertain parameters¹². The order of the polynomials, p , is a user choice – it requires that the sampling of uncertain variables in each direction is sufficient to provide enough points. When the PCE has converged, higher order coefficients should decay to zero and increasing p further causes no further change to the results. An illustration of how the PCE maps input variable uncertainty distributions to output variable distributions is shown in Figure 4.8.

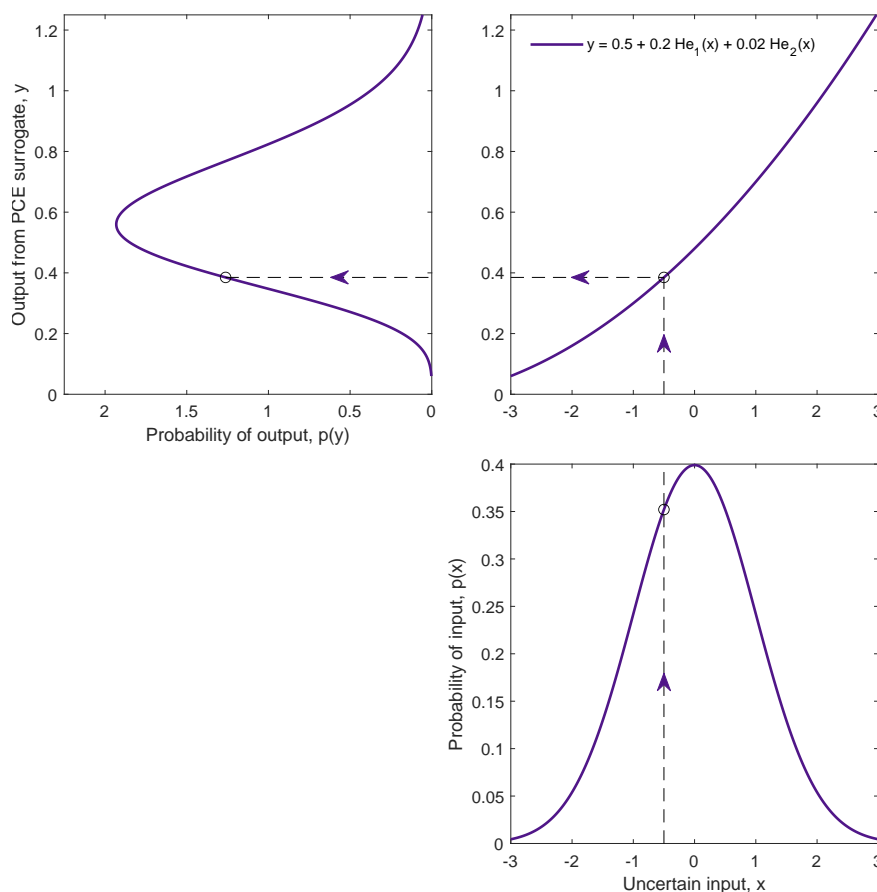


Figure 4.8: Example of a PCE, where a standard normal distribution of a single variable, x , is mapped through 0, 1 and 2 order Hermite polynomials to produce the PDF of an output y . Adapted from the example given by Debusschere (2017).

The first step of PCE defines the orthogonal polynomials ϕ_n . The form of the orthogonal polynomials is linked to the form of the PDF of the uncertainty parameter of interest. For example, if the

¹² This expression is for the ‘total-order expansion’ method of creating a PCE, where each parameter uses the same polynomial order, p . Other options are available that can reduce the number of terms, including allowing each parameter to select its optimal order (Dalbey *et al.*, 2020).

Mathematical Methods for Propagating Uncertainty

uncertainty in a variable were normally distributed, the orthogonal polynomials will be Hermite polynomials; for a uniformly distributed input, they will be Legendre polynomials (Dalbey *et al.*, 2020). In practice, software toolboxes can choose and generate these orthogonal polynomials internally.

The next step estimates the expansion coefficients c_n . These will often be derived numerically, either:

- Using regression to minimise the difference between the value predicted by the polynomial expansion and the full-order model output results, corresponding to input samples chosen by a suitable experimental design method (such as LHS), or
- Using a ‘spectral projection’ method, exploiting orthogonality in the sampling (or sparse grid sampling) of the parameter space.

Having the expansion coefficients results in an analytical expression, from which the statistical moments defining the uncertainty in the output (such as the mean and variance) can be calculated directly. In addition, the polynomial chaos expansion, once derived, can be interrogated like a response surface. PCE is an attractive surrogate model method because it requires a relatively small number of full-order model samples to generate. However, its implementation is not straightforward¹³ and the number of polynomials involved can become large, so it is recommended that a software toolkit such as Dakota¹⁴ or Uncertainpy (Tennøe *et al.*, 2018) is used to exploit this method.

PCE can be much faster (by requiring fewer full-model evaluations) than other uncertainty quantification methods, as long as the number of uncertain parameters is relatively low and the inputs are uncorrelated (Roy and Oberkampf, 2011). However, as the number of uncertain parameters increases, so does the number of full-order model evaluations required to achieve the same convergence.

4.4.2.5 Kriging model

Kriging based surrogate models, also known as Gaussian process regression, are a form of response surface. As with other types of surrogate, a Kriging model can be interrogated many times to propagate the uncertainty in model parameters through to model outputs.

A Kriging model is formed of two parts:

- A regression of the samples from the full-order model.
- The Gaussian process term.

The Gaussian process term takes into account the correlation between samples to estimate the uncertainty in the regression. Kriging methods derive this uncertainty on the basis that samples taken close together in parameter space will be highly correlated whereas those taken further apart are less likely to be highly correlated. Therefore, a Kriging model can provide not only a prediction of the model response, but also the uncertainty in the prediction as a result of using the surrogate model.

¹³ It is noted that PCE can be implemented both intrusively (through direct modifications to the source code) and non-intrusively (Najm, 2009). The focus here is on non-intrusive methods, because it is assumed that an analyst will either not have access to the source code of the full-order model, or sufficient time or knowledge to modify it.

¹⁴ dakota.sandia.gov

This is illustrated in Figure 4.9. It can be seen that close to the training data, the uncertainty in the prediction is small, but this uncertainty grows as the surrogate model is sampled further from the training dataset.

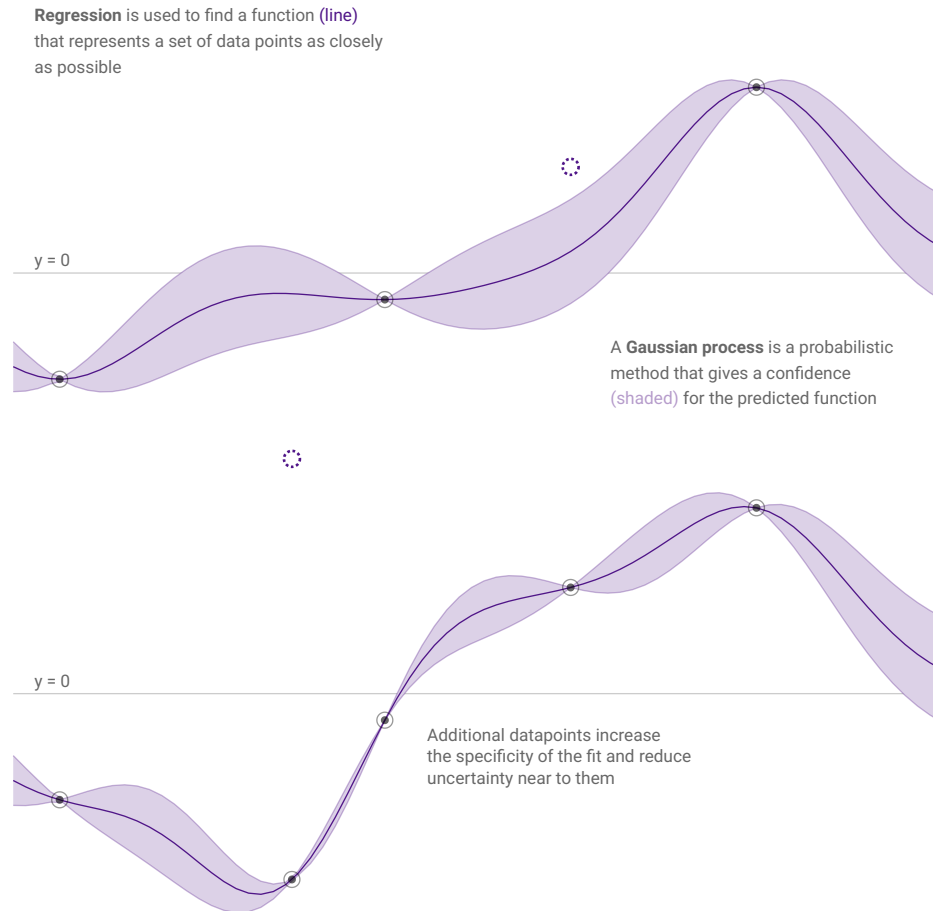


Figure 4.9: Example of a Kriging model, showing the uncertainty bounds, and that they reduce close to the sampling points used to construct it. Adapted from Görtler *et al.* (2019).

Kriging interpolation techniques were originally developed for use in the field of geostatistics, where geological samples are taken at irregularly spaced intervals and Kriging interpolation is used to create underground maps based on this information. They are also widely used in machine learning (Rasmussen and Williams, 2006) as well as in nuclear applications (Banyay *et al.*, 2019). Kriging methods have the advantage that they can accommodate irregularly spaced data and capture 'peaks' and 'troughs' in the full-order model, thereby allowing for models with more complicated structure in responses.

It is noted that Kriging models can become unstable as the number of sample points from the full-order model increases. This is because the correlation matrix that forms part of their description can become ill-conditioned as the distance between sample points decreases (Giunta *et al.*, 2006). Therefore they are best suited to sparse sample sets.

Mathematical Methods for Propagating Uncertainty

4.4.2.6 Model Order Reduction

ROMs take the full-order, high-fidelity model and project this onto a lower-dimensional basis which is more efficient and faster to solve. Reduced order methods have the advantage that they retain the underlying structure of the model, however they can be more intrusive than some of the other surrogate model methods discussed in this section. This means that the full-order model cannot necessarily be treated as a 'black box', and requires knowledge of the underlying model form in order to derive the reduced order model.

POD, balanced truncation and Krylov subspace methods are examples of methods used for Projection-based ROMs (Benner *et al.*, 2015). POD (also known as the Karhunen-Loève method in the field of stochastic theory, or principal component analysis in the field of statistical analysis) is a snapshot-based ROM method that has been applied to CFD problems (Section 3.4.4).

In POD, a 'snapshot matrix' is generated, which is a collection of observations of the quantities of interest (such as the velocity or temperature fields) at either a series of instants in time, or as separate realisations¹⁵ of a flow, but with a parametric variation (in flow rate, for example). From this snapshot matrix, a set of empirical eigenfunctions are derived which form the basis of the ROM. These eigenfunctions identify the most energetic modes of the full-order model. The ROM is then obtained from the Galerkin projection of the governing equations onto this POD basis. Georgaka *et al.* (2020) provides a more detailed description of the derivation of a ROM via POD.

4.5 Sensitivity Analysis

Sensitivity analysis is used to determine which model inputs have an important effect on the output metrics. Although similar to uncertainty quantification, the focus of sensitivity analysis is on identifying the important variables and relationships in a model, rather than in numerically quantifying and propagating uncertainty through to the model output. Local sensitivity analysis involves studying the relative influence of different parameters at a given point in the input space. On the other hand, global sensitivity analysis examines the effect of one parameter whilst other parameters are also varied, thus allowing the effect of interactions between variables to be accounted for. In addition, the analysis is not sensitive to a chosen nominal point in input space (Kucherenko and Iooss, 2017). Global sensitivity analysis techniques are the focus of the remainder of this section.

Sensitivity analysis usually involves the following steps:

- Identify the model responses of interest (the key safety or performance metrics).
- Identify potentially important input parameters.
- Propose plausible bounds on these input parameters for the sensitivity analysis.
- Use experimental design techniques (Section 4.3) to construct a dataset on which to perform the sensitivity analysis.
- Execute the full-order (or surrogate model) model evaluations for the chosen dataset.
- Perform sensitivity analysis by calculating or interpreting the measures and metrics discussed below. Many software packages have built-in functionality that can assist with this.

¹⁵ Chosen using the techniques described in Section 4.3.

Mathematical Methods for Propagating Uncertainty

4.5.1 Scatter Plots

While not a ‘measure’ of sensitivity per-se, scatter plots provide a useful first step to examine the impact of different model inputs on the model outputs of interest, allowing a visual identification of trends and structure in the variation of model output with variations in input.

4.5.2 Sensitivity Coefficients

The Jacobian is the first order gradient (linear sensitivity coefficient) of the variation of each of m model outputs, f , in response to n model inputs, x :

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Where the mathematical form of the model equations is suitable and sufficiently simple, then analytical methods are available such that the sensitivity and uncertainty in them can be derived without needing to apply numerical and statistical methods. In this case, and where there is only one output variable, a Taylor expansion can be used to determine the sensitivity at higher orders, and then be used to analytically propagate uncertainty. For mathematically analysable models with multiple input and outputs, partially differentiating the expressions provides the Jacobian algebraically, which can then be evaluated at any point in the parameter space. However, the Jacobian is typically estimated using numerical methods, because the model used is too complex to assess analytically. A particular numerical Jacobian only applies at one point in parameter space (such as the centre of the input distributions, so is a local sensitivity).

It may be beneficial to deliberately frame the complexity and mathematical representation of a system to facilitate the efficient application of analytical methods, including more sophisticated techniques such as adjoint methods (Cacuci, 2003).

4.5.3 Correlation Coefficients

Correlation coefficients are a method of quantifying the statistical relationship between two parameters or two sets of data. There are two commonly used correlation coefficients (Hofer, 2018):

The simple correlation coefficient (also known as the Pearson correlation coefficient) measures the strength and direction of a *linear* relationship between two variables. The coefficient can vary between -1 and 1, with a correlation of 1 indicating a perfect, positive correlation and -1 indicating a perfect, negative correlation. If there is no or little linear correlation, the coefficient will be close to 0.

The rank correlation coefficient (also known as the Spearman coefficient) performs correlations on ranked data giving a measure of the *monotonicity* of the relationship. In order to obtain ranked data, it is ordered such that the smallest value in a set of samples is given the rank 1, the next smallest the rank 2 and so on. Rank correlation coefficients can be used if there is a non-linear relationship between the two variables.

Mathematical Methods for Propagating Uncertainty

Each type can have a 'partial' variant

Partial correlation coefficients are similar to the simple and rank correlation coefficient, but allow for the correlation between two variables to be determined whilst controlling for the effects of the other variables present in the problem.

It is also possible to determine the statistical significance of the correlation. This can be expressed as a p-value (between 0 and 1), which is often returned as an additional output with the correlation coefficient when using a statistical toolbox.

4.5.4 Sobol Indices

Whilst correlation coefficients test for a linear or monotonic relationship, Sobol indices (also known as variance-based decomposition) provide a measure of how much the output variance is attributable to each input variable.

There are two principal sensitivity indices. The first is the main effect sensitivity index, S_i (Saltelli *et al.*, 2007, Hofer, 2018):

$$S_i = \frac{\text{Var}(E(f|x_i))}{\text{Var}(f)}$$

where E is the 'expected' value (the mean) and Var is the variance. The numerator¹⁶ is the output variance of the main effect of x_i , attributable to that parameter only. The denominator is the total variance in the output. Higher order terms can also be calculated, known as interaction terms. It is required that

$$\sum_i S_i + \sum_i \sum_{j>i} S_{ij} + \sum_i \sum_{j>i} \sum_{k>j} S_{ijk} + \dots = 1$$

A large main effect index indicates that the input parameter in question has an important effect on the output of interest.

The second metric is the total effect index T_i :

$$T_i = \frac{E(\text{Var}(f|x_{-i}))}{\text{Var}(f)} = \frac{\text{Var}(f) - \text{Var}(E(f|x_{-i}))}{\text{Var}(f)}$$

This measures the contribution to the output variance of x_i , including the effects of its interactions with other variables¹⁷, and has the advantage that it does not need to calculate the interaction terms (S_{ij} and S_{ijk}) separately. A larger total effect index for a given parameter compared to S_i indicates that this parameter interacts with other parameters, and the sum of total effect indices is greater than or equal to one (it only equals one when there are no interactions, Glen and Isaacs, 2012).

Although analytically calculating the Sobol indices is possible, the number of sensitivity indices quickly becomes unmanageable as the number of inputs increases. This means that usually the sensitivity indices are solved for numerically. One exception to this is if a PCE surrogate model has been created – the properties of the PCE mean that the Sobol indices can be calculated analytically (Sudret, 2008).

¹⁶ The vertical line in $E(f|x_i)$ denotes the 'conditional expectation', e.g. the mean of the function f , as it depends on x_i .

¹⁷ $(f|x_{-i})$ denotes being conditional on all inputs except x_i .

4.6 Uncertainty Quantification by Propagation

Quantification of the effect of uncertain model parameters and of initial and boundary conditions is a common application for UQ. This takes the uncertain parameters, characterised into probability distributions using the methods described above, and propagates them through a modelling tool to predict the distribution of the model outputs. This section covers the propagation and UQ where all input variables are characterised by PDFs. How to combine these methods with other forms of uncertainty is discussed in Section 4.8.

Sampling approaches are the simplest methods of parametric uncertainty propagation, so are usually the first step in any UQ process:

Sample from the assumed distribution for each input parameter: LHS should be the default choice of sampling in most cases. It is important to have a sufficient sample size to ensure that there is adequate sampling from all areas of a parameter's probability distribution; Swiler and Giunta (2007) suggest that the number of samples should be at least 10 times the number of uncertain variables.

Propagate the sampled values through the model: Evaluate the simulation results for each set of sampled values, treating the numerical model as a 'black box'. Because all samples are independent and known at the outset of the calculation, evaluations can be carried out in parallel to speed up the analysis.

Interpret the model output: For each output metric, the evaluation results can be plotted as a histogram to assess the statistics and distribution of the predicted values. The median of this distribution represents the best estimate value and its width (calculated as the variance, σ^2) represents the uncertainty. The cumulative sum of the histogram provides an estimate of the CDF, which can be used, for example, to determine the likelihood of output metric exceeding a limit value. Taking more samples will generally allow an output distribution to be smoother, its median and variance to be known with more certainty (this is discussed further in Section 4.7), and the lower probability ends of distribution to be better resolved.

Guidance on what tools can be used to help perform these steps, and suggested visualisation methods suitable for interpreting the outcome are discussed in Section 2.2.

4.6.1 Use of Surrogate Models

If the model to be sampled is fast enough to evaluate, then it can be used directly to gather enough data to provide the necessary statistics and output distribution. However, if the model is resource intensive, such that each sample evaluation requires extensive computation, it may not be practical to apply sampling UQ to it directly.

In this case, surrogate models (Section 4.4) can be used. They can be used for purely computational efficiency reasons, where a relatively small number ($O(10^2)$) of sampled evaluations of the 'full-order' model can be used to construct the surrogate, which can itself be sampled and rapidly evaluated a much larger number of times ($O(10^4)$ to $O(10^5)$) to provide the output statistics and distribution. They can also be employed as an efficient way to provide more refined and additional information about the results.

Mathematical Methods for Propagating Uncertainty

There are a number of considerations when applying a surrogate for UQ:

- Giunta *et al.* (2006) discuss whether, when and why to use a surrogate model (referring to them as trend models):

... this study attempts to answer the following question, 'Given a set of N simulation output data samples, is it more accurate to estimate UQ statistics directly from the N data points, or is it more accurate to extract trend information from the N data points, and then estimate UQ statistics from the trend models?'

Although it only deals with a simple problem with a small number of uncertain inputs, the decision about using a surrogate model becomes implementation and case specific, so generalised guidance is not possible. They conclude that caution is needed, and suggest applying several surrogate model approaches, checking that they provide a similar result, checking for a strong dependence on N , and comparing between a surrogate method and a purely sampled result to select a suitable technique. They also found that for small N , there was not necessarily a benefit obtained by using a surrogate.

- It should be remembered that using a simplified model to propagate the uncertainties introduces its own contribution to the uncertainty, and this should be able to be estimated and shown to be sufficiently small. Kriging models can provide this uncertainty estimate as part of their construction
- Fitting Kriging models can become unreliable with larger numbers of samples (Giunta *et al.*, 2006), so a higher N is not necessarily better.
- PCE needs the model response to be smooth, but has the significant benefit that it can provide Sobol indices and the mean and variance of the response directly, without sampling them further. They can, however, also be rapidly sampled to provide a well resolved CDF.

4.6.2 Reliability Methods

In some situations, the mean behaviour of a system is not of interest, only the margin to, or behaviour near to a limit. In a structural integrity assessment, this is known as 'reliability', and is concerned with the approach to a 'cliff-edge' failure condition. Therefore, understanding the whole span of the CDF of a response with equal resolution and accuracy is not desired. Uncertainty propagation via reliability methods focuses on finding and characterising the parameter space near to the failure limit, and evaluating the likelihood of exceeding it based on the uncertainty in the input parameters.

Traditional First Order Reliability Method (FORM) and Second Order Reliability Method (SORM) approaches approximate a linear or quadratic failure surface in parameter space, and determine the probability of being on either side of it to be assessed. There are several more advanced variants available to find the failure surfaces, that are similar to optimisation searches – they adaptively select new evaluation points to resolve the failure region using 'importance sampling' (MacKay, 2003). Reliability methods are therefore often more efficient at resolving the 'tails' of input distributions for low probability conditions than sampling methods. These methods often make use of underlying Gaussian process (Kriging) surrogate models in their optimisation evaluations.

Further discussion and guidance on choosing an uncertainty quantification approach is available

from the documentation of tools used to perform it, such as Dakota (Adams *et al.*, 2020, Chapter 5).

4.7 Assessment of Output Uncertainty from Propagation

The aim of the uncertainty quantification process of propagating uncertain inputs defined by a probability distribution is to calculate the PDF or CDF of an output metric. This section outlines some of the principles useful for interpreting the probability distribution representing the uncertainty in the output metric.

4.7.1 Confidence Intervals and Levels

The principle of a Confidence Interval (CI) is illustrated in Figure 4.10, where histograms of an output metric from simulation results visualises the effect of aleatory uncertainty in the inputs. A CI represents the range of values within which the output metric of interest is potentially contained. A confidence interval is stated with specified confidence level (sometimes called the degree of confidence). The distinction between confidence and probability, and the correct statistical interpretation of CIs can be subtle and it is hard to express the details concisely (DeGroot and Schervish, 2012). Following the explanation of Navidi (2020), consider a population with a mean, μ , from which a finite number of samples is taken, giving a sample mean, \bar{X} . 95 times out 100, when taking such a random sample, the (true, but unknown) population mean will be expected to lie within the 95 % CI of the sample mean.

Two types of confidence interval exist – one-sided and two sided. Figure 4.10 demonstrates a two-sided confidence interval (a, b) , which brackets the uncertainty above and below the central-estimate value. A one-sided confidence interval brackets all of the uncertainty distribution on one side of a value: (a, ∞) is the lower one-sided interval, and $(-\infty, b)$ is the upper one-sided interval. In the example demonstrated in Figure 4.10, the upper end of the 95 % two-sided confidence interval is equivalent to the one-sided upper 97.5 % confidence interval. One-sided intervals are used in failure assessments where only a lower (or upper) bound of a limit is of interest. A two-sided interval gives an indication of the spread of the uncertainty in a quantity.

The CI represents a trade-off – for a given distribution, narrower limits can be stated on the value of the central estimate, but they then must represent a lower confidence level.

For a symmetric uncertainty distribution, the central-estimate (median or 50th percentile) is equivalent to the mean and mode. However, if the calculated uncertainty distribution is asymmetric then this no longer holds. The mode is the most likely occurring value, at the peak of the probability distribution, the mean is the average value.

In a forward propagation of uncertainty calculation, the probability density of the output metric can be found by evaluating a set of Monte-Carlo samples (often using a surrogate model). The result of this approach is an empirical distribution of values of the output metric. The empirical distribution of Monte-Carlo samples represents the calculated uncertainty on the output metric. The median value and confidence interval can be found by either taking relevant percentiles of the empirical distribution (e.g. 50th, 2.5th and 97.5th for a 95 % CI)

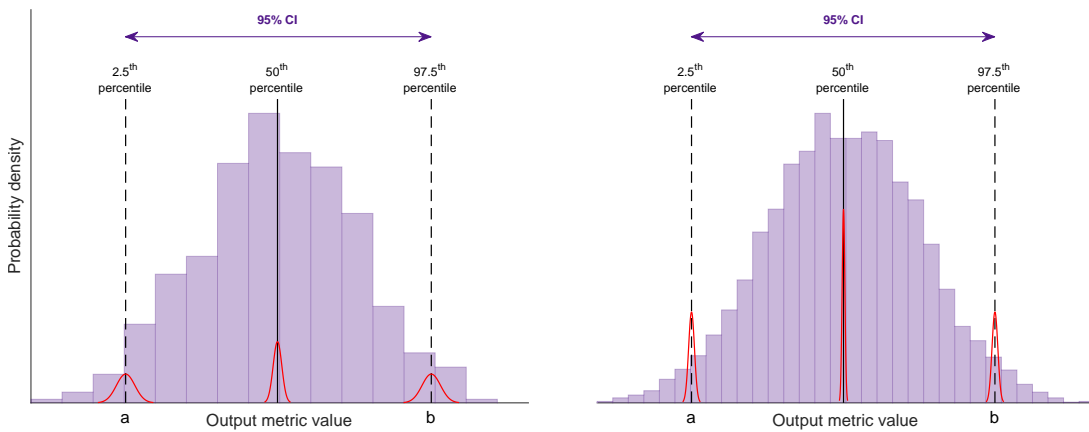


Figure 4.10: Example of the 95 % CI (a , b) of an output metric (represented by a normal distribution randomly sampled either 500 times (left) or 5000 times (right) to simulate a range of modelling results). Also indicated is how the percentiles of the sampled uncertainty distribution are also uncertain in their own right – the uncertainty in them that arises from the finite number of samples can be reduced with an increasing the number of evaluations (although this is not the only source of uncertainty).

A concept that often occurs when introducing UQ to an analysis process is that of a ‘baseline’ model that may have previously served as a deterministic model, and serves as a starting point around which uncertainty analysis is built. If symmetric uncertainty distributions are placed around its inputs, and the process modelled has an approximately linear response, then the results of the baseline model will lie near to the central-estimate (or 50th percentile) or the mode of the output uncertainty distribution. However, this does not always need to be the case. It may be that the chosen uncertainty distributions of the inputs do not have the baseline model at their centre when they are assessed, or they are not symmetric and the model is non-linear, such that the effect of the dominant uncertainty is to influence the outputs to be mainly in one direction. It is therefore not inherently expected or required that the baseline results are representative of the median, after applying UQ.

4.7.1.1 Confidence Interval Uncertainty

It should be noted that the confidence intervals derived from any uncertainty assessment calculation will themselves contain uncertainty. This can originate from two sources:

- The numerical calculations in the uncertainty assessment are not exact. This could be related to performing a limited number of model calculations, limitations in the experimental design or uncertainty in a surrogate model. In general this uncertainty can be quantified and should be considered when interpreting a quoted CI and confidence level. It is more significant when considering an output metric of interest in the tails of the uncertainty distribution when compared to the mean (i.e. more samples will be required to determine the 99th percentile to the same level of accuracy as the mean). This is shown in Figure 4.10, where the uncertainty on different percentiles of an underlying distribution are indicated.
- The formulation of the uncertainty assessment is not definitive. For example, approximations, assumptions and expert judgements made during derivation of the uncertain input distributions, or source of uncertainty not considered in the analysis. In general, this uncertainty can

be difficult to quantify.

These items need to be considered when interpreting the results of the uncertainty quantification in the context of overall nuclear safety.

4.7.2 Statistical Tolerance Limits

Where confidence limits are the values within which a given population parameter, such as the mean, is expected to lie, statistical tolerance limits are the values within which a stated proportion of the population is expected to lie (NIST, 2021).

The GRS method is a propagation method based on Wilks' formula for sampling and tolerance/confidence levels (Wilks, 1942). While it is a UQ method in itself ...

A significant advantage of this methodology is that no a priori reduction in the number of uncertain input parameters by expert judgement or screening calculations is necessary to limit the calculation effort. All potentially important parameters may be included in the uncertainty analysis. The method accounts for the combined influence of all identified input uncertainties on the results. This would be difficult or even impossible to achieve by a priori expert judgement of loss of coolant accidents or transients. The number of calculations needed is independent of the number of uncertain parameters accounted for in the analysis. It does, however, depend on the requested tolerance limits, that is, the requested probability coverage (percentile) of the combined effect of the quantified uncertainties, and on the requested confidence level of the code results. The tolerance limits can be used for quantitative statements about margins to acceptance criteria.

Glaeser, 2008

... the key point is about the definition and interpretation of a result given by computing ($u\%$, $v\%$) as statistical tolerance limits (Hofer, 2018):

- *The two-sided ($u\%$, $v\%$) statistical tolerance limit contains at least $u\%$ of the population of possibly true answers to the assessment question at a confidence level of at least $v\%$.*
- *The upper [one-sided] ($u\%$, $v\%$) statistical tolerance limit says that at least $u\%$ of the population of possibly true answers to the assessment question do not exceed this limit value at a confidence level of at least $v\%$.*

The confidence level ($v\%$) is specified because the statements about the quantile covered ($u\%$) are obtained from a random sample of limited size (illustrated in Figure 4.10). The attraction of these limits are that the number of simulations required to satisfy them can be determined in advance, and is independent of the number of uncertain variables. For example, $N = 59$ model evaluations are needed for a one-sided (95%, 95%) tolerance limit, and $N = 93$ are needed for a two-sided (95%, 95%) tolerance limit (Glaeser, 2008). Reduced allowable confidence permits a lower number of evaluations, where $N = 45$ is needed for a one-sided (95%, 90%) tolerance limit.

While this predictability and independence from the number of uncertain variables offers benefits, and does not require the application of surrogate models to the samples, the confidence limits are not automatically achieved simply by performing any selection of N samples. The validity of the

Mathematical Methods for Propagating Uncertainty

tolerance limits come with other requirements and limitations (Hofer, 1999, Hofer, 2018, Porter, 2019), such as being non-tolerant of code run failures, requiring low model and numerical uncertainty and placing requirements on the quality and completeness of the definition of the input uncertainties. It is also restricted to using simple random sampling – methods such as LHS cannot be used.

4.7.3 Chebyshev's Rule

Chebyshev's rule states that for any probability distribution for which a mean and standard deviation can be derived (Vose, 2008):

For any number k greater than 1, at least $(1 - 1/k^2)$ of the measurements will fall within k standard deviations of the mean.

For example, for any probability distribution 75 % of the measurements must fall within 2 standard deviations of the mean (± 1). Chebyshev's rule is a more general version of the rule that for, a normal distribution, 68 %, 95 % and 99.7 % of the measurements must fall within 2, 4 and 6 standard deviations of the mean (also known as the 68-95-99.7 rule). By necessity, Chebyshev's rule is conservative, and a more precise estimate can be made if the specific distribution form is known. However, Chebyshev's rule may be used to define:

1. A conservative interval range when characterising an uncertain epistemic input.
2. A confidence interval on an output metric when the precise distribution form is unknown or uncertain.

4.8 Combining Uncertainty

The propagation of uncertainty discussed in Section 4.6 is associated with input uncertainties that can largely be described by continuous probability distributions. However, as noted in Section 4.2, epistemic uncertainties may be represented by intervals and combined uncertainties via imprecise probability. This section discusses methods and techniques by which these and other forms of uncertainty can also be included in a UQ calculation. The propagation and combination of different sources of uncertainty, and the mathematical tools to reduce the computational cost of such activities, is an area of current research (for example, Schöbi and Sudret, 2017).

4.8.1 Propagating Epistemic and Aleatory Uncertainties

A common approach taken to propagate and combine epistemic and aleatory uncertainties is to use nested sampling. In an outer sampling loop a sample is taken for each of the epistemic uncertainties, giving a realisation of a single possible epistemic uncertain position. In the inner loop, for the given epistemic realisation, a probabilistic calculation of the aleatory uncertainties, as described in Section 4.6 is performed. The result of this nested sampling is an ensemble of aleatory distributions of output metric uncertainty, similar to that shown in Figure 4.3.

The epistemic uncertainty may be represented either by a probability distribution (based on the belief of an analyst or other numerical evidence) or an interval. These different interpretations

Mathematical Methods for Propagating Uncertainty

of the epistemic uncertainties give rise to two common methods of implementing the outer loop described above.

4.8.1.1 Second-order Monte-Carlo

In second-order (or nested) Monte-Carlo calculations, the epistemic uncertainties in the outer-loop are represented by probability distributions similar to the inner loop described in Section 4.6. Often these distributions will be subjective in nature if there is only limited information available to describe the uncertainty. The output of this method is that each CDF generated has a probability of occurrence. The overall uncertainty distribution is the convolution of the CDFs taking into account their relative probability of occurrence.

4.8.1.2 Probability Bounds Analysis

The epistemic uncertainties alternatively can be represented by an interval, about which there is no knowledge of the structure of the uncertainty. In this case, the outer loop consists of taking samples from the epistemic uncertain intervals with no indication of the underlying probability. The output is similar to that for second-order Monte-Carlo, but the individual CDFs have no knowledge of their underlying likelihood, and it is appropriate to use a p-box representation of the output metric uncertainty using a PBA. The combined output metric uncertainty is defined by the upper and lower bound CDFs of the resulting p-box, so it is necessary to determine the combination of values of epistemic intervals that produce the widest output range. Defining these bounding cases is not trivial for a multi-dimensional problem and Roy and Oberkampf (2011) describe how it can be formulated as a constrained optimisation problem.

The resulting p-box of the output metric gives a visualisation of the relative importance of the epistemic and aleatory uncertainty on the output metric and can be used to target future work to investigate the most significant specific areas of reducible epistemic uncertainty. For example, performing sensitivity studies investigating the impact of the epistemic uncertainty interval ranges the output metric, the p-box can be 'pinched' inwards to determine if further work to reduce uncertainties can have a meaningful impact on the overall output metric uncertainty. DST may also be used in a similar outer loop but instead of a single interval defining the epistemic uncertain input, the belief and plausibility functions are used instead.

4.8.2 Model Uncertainty

Uncertainty in a thermal hydraulic model (model uncertainty) is harder to quantify than uncertainty in input values. It is inevitable that model uncertainty will depend on a modelling decision or a subjective expert opinion, especially when a model is used to predict the consequences of a hypothesised fault where many phenomena interact, and that these epistemic uncertainties cannot be easily characterised and propagated in the UQ. This is especially true when there is little or no experimental data, so it is not possible to obtain validation evidence for the situation being modelled (such as a severe accident).

The approaches discussed in this section may significantly increase the number of calculations required to quantify the uncertainty in an output metric. As a result, these methods may be best applied to system models (Section 3.3) which are less computationally expensive than more com-

Mathematical Methods for Propagating Uncertainty

plex CFD models. Furthermore, system code modelling approaches may require a greater degree of approximation and modelling judgement (introducing user effects) than a CFD model and hence the epistemic uncertainties on the model may be more important on the output metric and require greater consideration.

In practice epistemic uncertainty in the model form may be significant and difficult to quantify in any meaningful way. The effort expended in characterising and propagating this depends on how important it is to understand its effects in context of the overall use of the model output.

Model uncertainty may be formally incorporated into the UQ calculation by one of the following three methods (Tucker and Ferson, 2003):

1. Try to re-express the model uncertainty as a probabilistic or epistemic uncertainty in the UQ calculation. This way it can be propagated through model as if it were another form of uncertainty, characterised using the methods described in Section 4.2 and propagated through the UQ using the methods outlined in Section 4.8.1. One possible approach is to incorporate into the underlying model a meta-parameter which represents an average over many possible different model uncertainties. The approach requires a subjective view of the probability associated with the parameter to represent model uncertainty and the physical reasonableness of this approach can be questioned. However, once implemented, the methods described above can be used to quantify the impact on the output metric uncertainty. For example, this could be done by characterising model uncertainty via an epistemic interval meta-parameter in the outer sampling loop. As a less ideal, but potentially more practical option, the meta-parameter could be assumed to be a continuous distribution and incorporated into the inner sampling loop. However, this may make it difficult to separate the contributions of epistemic and aleatory uncertainties in output metric.
2. Work-through each modelling decision, assumption or approximation using a set of structured sensitivity studies to determine the impact on the output metric of interest. In Section 4.6 the model being used for propagation was treated as a 'black box', where the model input parameters were perturbed and the effect on the outputs were observed without any interrogation of the internal operations of model itself. In 'white box' modelling, the user 'peeks inside the box', focusing specifically on internal knowledge of the code to assess model uncertainty. However, this rapidly becomes difficult for any complex model and performing this analysis for any more than a few key cases is unlikely to be practicable.
3. Try to bound or judge the range of possible outcomes from any credible model which could exist and use this to define the epistemic uncertainty on the model output. This uncertainty may then be appended to the p-box defined by a PBA propagation of uncertainty. An assessment of the magnitude of this uncertainty may be based on evidence from model validation and extrapolation analysis (Roy and Oberkampf, 2011). In practice, this may be difficult for complex models which contain a number of poorly understood phenomenon and the relationship between the output metric of interest and available validation data is unclear.

Finally, model uncertainty may also incorporate aspects of numerical uncertainty in the UQ calculation route. The most common sources to consider are likely to be (Mahadevan and Sarkar, 2009):

- Discretisation or convergence uncertainties in the underlying physical model.

Mathematical Methods for Propagating Uncertainty

- Errors in the UQ analysis, e.g. due to performing a limited number of Monte-Carlo samples and uncertainties associated with using a surrogate model (e.g. the surrogate model residuals).

These can most readily be included in the UQ by incorporating them in the outer Monte-Carlo sampling loop, or in a PBA by appending the uncertainties to the edges of the p-boxes derived from propagating the other epistemic and aleatory uncertainties.

4.8.3 Impact on the Number of Model Calculations Required

Both second-order Monte-Carlo and PBA can substantially increase the computational requirements of the UQ. For example, Roy and Oberkampf (2011) recommend that a minimum of three LHS samples are taken for each epistemic uncertainty, in combination with all of the remaining epistemic uncertainties. The impact of this recommendation is that the minimum number of samples required increases as the cube of the number of epistemic uncertainties. Given that this represents the outer loop in the nested Monte-Carlo sampling, the aleatory uncertainties must also be propagated through the model for each combination of epistemic uncertainties.

This issue is often known as the ‘curse of dimensionality’ where the number of samples required rapidly becomes computationally limiting, for example see Kreinovich *et al.* (2006). The use of sampling methods and surrogate models discussed in Sections 4.3 and 4.4 can help mitigate this, but for more than a small number of epistemic uncertainties it can be a significant limitation.

When embarking on a UQ calculation, consideration needs to be given to the timescales and computational requirements of the task compared to the scope that can be achieved using the methods described in this technical volume. This should include the topics outlined in Section 4.3 along with the number of epistemic uncertainties that need to be propagated through the calculation.

4.8.4 Inter-model Comparison

It is possible for multiple analysts or teams to develop diverse models of the same NTH phenomenon (diverse in terms of underlying code, modelling assumptions and/or solution method). These models will not give identical results, even when using the same underlying physical model, due to user uncertainty. As a result, combining the results from a number of such models can be used to assess the epistemic model uncertainty via the inter-model spread (this is used in disciplines where the underlying physical models are not fully understood, such as climate modelling IPCC, 2014).

The cost and resource requirements needed often make it impractical to develop multiple models in all circumstances, but for calculations with potentially large uncertainties and significant consequences, consideration should be given for developing a small number of separate and independent models to examine the epistemic uncertainties on model formulation and user effects. Examining the output of such models, along with an estimation of the uncertainty on the output metrics can be highly informative and build confidence in the overall conclusions of the modelling work, and this is the basis for international benchmarking activities, as discussed in Section 3.

Mathematical Methods for Propagating Uncertainty

4.8.5 Presenting Overall Uncertainty in an Output Metric

To make best use of the efforts applied in performing UQ, the combined overall uncertainty in the outputs of a model needs to be summarised and presented to relevant stakeholders and considered in the context of the application and importance of the predictions, especially if they relate to nuclear safety. The calculated uncertainty on an output metric could be used in several ways:

- The result could be compared against a performance target (or safety limit), as shown in Figure 1.2, to provide substantiation of acceptability of the operation of the system. For example, demonstrating robustly that fuel cladding remains intact during a specific fault scenario even when accounting for modelling uncertainties. For a nested Monte-Carlo approach, the output uncertainty will be a PDF or CDF from which the confidence level that performance target is reached can be calculated directly. For a PBA the output will be a p-box against which the performance target can be compared.
- The result could be used to set a performance target (or safety limit) for other aspects of the system, which can be used to ensure safe operation. For example, a calculation for the temperature of spent fuel and the uncertainty on this value could be used to constrain the timescales of certain fuel handling operations, or the uncertainty on time to fuel damage could be used to inform the required response time of automated plant protection C&I systems.
- The result could be used to identify key sources of epistemic uncertainty which must be further reduced (through further experimental measurements, for example) in order to substantiate future safe performance of a system. For example, a current safety limit may become increasingly important as plant degrades during operation. A UQ and SA prediction of the core age when the safety limit is reached could be used to define an experimental programme or modelling activities to reduce epistemic uncertainties and inform life-extension or to help plan decommissioning activities.

Most importantly, consideration must be given to the nuclear safety consequences associated with a calculation. For some analyses, it will be desirable to demonstrate that there is additional margin above the calculated uncertainty, which implies that reaching the safety limit is essentially impossible (there remains substantial margin between the upper limit of the calculated uncertainty and the safety limit). This could be required for assessments for which a failure might result in a significant level of radiological release. In other assessments, it may be more appropriate to demonstrate that, to a given confidence level, the probability of failure is below a certain threshold.

It is also possible that there is uncertainty on a performance target which also needs consideration alongside uncertainty in the output metric. For example, a fuel damage temperature may be derived from experiments which contain their own uncertainty. Figure 4.11 demonstrates this scenario. Under such circumstances a variety of methods for interpreting the results of the UQ could be applied. For example, a conservative lower bound on the safety limit could be used; hypothesis testing could demonstrate that a percentile of the output metric distribution is less than a percentile on the performance target (to a given probability); or the Bhattacharyya distance (EMS, 2021) could be used to measure the level of overlap between the two distributions.

Mathematical Methods for Propagating Uncertainty

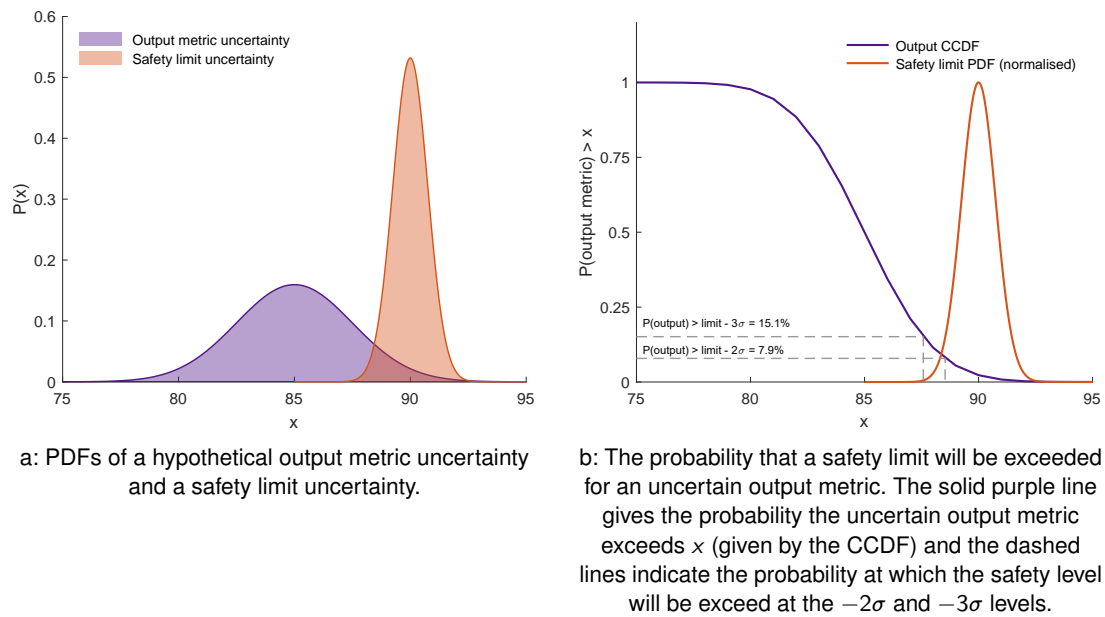


Figure 4.11: Example assessment of uncertainty in both an output metric and a safety limit.

The needs of a UQ assessment with respect to these issues should be derived from the requirements under which the modelling activity it supports is being commissioned, and can be structured within the organisational approaches discussed in Section 3.

5 References

- Acton M, Baglietto E** (2019) Addressing the Usage of CFD within the CSAU Framework for Nuclear Reactor Safety Analysis Simulations. In *Proceedings of 18th International Topical Meeting on Nuclear Reactor Thermal Hydraulics (NURETH-18)*.
- Adams B M, Bohnhoff W J, Dalbey K, Ebeida M S, Eddy J P, Eldred M S, Hooper R, Hough P D, et al.** (2020) Dakota A Multilevel Parallel Object-Oriented Framework for Design Optimization Parameter Estimation Uncertainty Quantification and Sensitivity Analysis: Version 6.12 User's Manual. SAND2020-5001, Sandia National Laboratories.
- Ahn S, Fessler J A** (2003) Standard Errors of Mean, Variance, and Standard Deviation Estimators. Technical Report 413, University of Michigan.
- Andreani M, Gaikwad A J, Ganju S, Gera B, Grigoryev S, Herranz L E, Huhtanen R, Kale V, et al.** (2019) Synthesis of a CFD Benchmark Exercise Based on a Test in the PANDA Facility Addressing the Stratification Erosion by a Vertical Jet in Presence of a Flow Obstruction. *Nuclear Engineering and Design*, 354, 110177, dx.doi.org/10.1016/j.nucengdes.2019.110177.
- Andres T H** (2002) Uncertainty Analysis Guide. AECL-12103, Atomic Energy of Canada Limited.
- ASME** (2009) Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer. ASME V&V 20-2009, American Society of Mechanical Engineers.
- Baccou J, Zhang J, Fillion P, Damblin G, Petrucci A, Mendizábal R, Reventos F, Skorek T, et al.** (2020) SAPIUM: A Generic Framework for a Practical and Transparent Quantification of Thermal-Hydraulic Code Model Input Uncertainty. *Nuclear Science and Engineering*, 194(8-9), 721–736, dx.doi.org/10.1080/00295639.2020.1759310.
- Banyay G A, Shields M D, Brigham J C** (2019) Efficient Global Sensitivity Analysis for Flow-Induced Vibration of a Nuclear Reactor Assembly Using Kriging Surrogates. *Nuclear Engineering and Design*, 341, 1–15, dx.doi.org/10.1016/j.nucengdes.2018.10.013.
- Banyay G A, Shields M D, Brigham J C** (2020) Efficient Global Sensitivity Analysis of Structural Vibration for a Nuclear Reactor System Subject to Nonstationary Loading. *Nuclear Engineering and Design*, 361, 110544, dx.doi.org/10.1016/j.nucengdes.2020.110544.
- Benner P, Gugercin S, Willcox K** (2015) A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems. *SIAM Review*, 57(4), 483–531, dx.doi.org/10.1137/130932715.
- Bhattacharyya B** (2018) A Critical Appraisal of Design of Experiments for Uncertainty Quantification. *Archives of Computational Methods in Engineering*, 25(3), 727–751, dx.doi.org/10.1007/s11831-017-9211-x.
- Bouhlef M A, Hwang J T, Bartoli N, Lafage R, Morlier J, Martins J R** (2019) A Python Surrogate Modeling Framework with Derivatives. *Advances in Engineering Software*, 135, 102662, dx.doi.org/10.1016/j.advengsoft.2019.03.005.

References

- Brooks F P** (1995) The Mythical Man-Month: Essays on Software Engineering. Anniversary edition, Addison-Wesley.
- Brown B B** (1967) Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts. RAND Corporation.
- Cacuci D G** (2003) Sensitivity and Uncertainty Analysis. Chapman & Hall/CRC Press.
- Chang Y H, Zhang L, Wang X, Yeh S T, Mak S, Sung C L, Jeff Wu C F, Yang V** (2019) Kernel-Smoothed Proper Orthogonal Decomposition–Based Emulation for Spatiotemporally Evolving Flow Dynamics Prediction. *AIAA Journal*, 57(12), 5269–5280, dx.doi.org/10.2514/1.J057803.
- Cho E, Cho M J, Eltinge J** (2005) The Variance of Sample Variance from a Finite Population. *International Journal of Pure and Applied Mathematics*, 21(3), 389.
- Choudhary A, Voyles I T, Roy C J, Oberkampf W L, Patil M** (2016) Probability Bounds Analysis Applied to the Sandia Verification and Validation Challenge Problem. *Journal of Verification, Validation and Uncertainty Quantification*, 1(1), 011003, dx.doi.org/10.1115/1.4031285.
- Coleman H W, Stern F** (1998) Closure to “Discussion of ‘Uncertainties and CFD Code Validation’” (1998, ASME J. Fluids Eng., 120, 635). *Journal of Fluids Engineering*, 120(3), 635–636, dx.doi.org/10.1115/1.2820714.
- Cooke R M** (1991) Experts in Uncertainty: Opinion and Subjective Probability in Science. Environmental Ethics and Science Policy Series, Oxford University Press.
- Cordier L, Bergmann M** (2008) Proper Orthogonal Decomposition: An Overview. Lecture series 2002-04, 2003-04 and 2008-01 on post-processing of experimental and numerical data, Von Karman Institute for Fluid Dynamics.
- Crawley M J** (2015) Statistics: An Introduction Using R. Second edition, Wiley.
- CSNI** (1994) User Effects on the Transient System Code Calculations. NEA/CSNI/R(94)35, OECD NEA Committee on the Safety of Nuclear Installations.
- CSNI** (1998) Report on the Uncertainty Methods Study. NEA/CSNI/R(97)35 Volume 1 and 2, OECD NEA Committee on the Safety of Nuclear Installations.
- CSNI** (2011a) BEMUSE Phase VI Report: Status Report on the Area, Classification of the Methods, Conclusions and Recommendations. NEA/CSNI/R(2011)4, OECD NEA Committee on the Safety of Nuclear Installations.
- CSNI** (2011b) Report of the OECD/NEA - Vattenfall T-Junction Benchmark Exercise. NEA/CSNI/R(2011)5, OECD NEA Committee on the Safety of Nuclear Installations.
- CSNI** (2013) Report of the OECD/NEA-KAERI Rod Bundle CFD Benchmark Exercise. NEA/CSNI/R(2013)5, OECD NEA Committee on the Safety of Nuclear Installations.
- CSNI** (2015) Best Practice Guidelines for the Use of CFD in Nuclear Reactor Safety Applications - Revision. NEA/CSNI/R(2014)11, OECD NEA Committee on the Safety of Nuclear Installations.
- CSNI** (2016a) The Nuclear Energy Agency/Paul Scherrer Institute Computational Fluid Dynamics Benchmark Exercise. NEA/CSNI/R(2016)2, OECD NEA Committee on the Safety of Nuclear Installations.

References

- CSNI** (2016b) Review of Uncertainty Methods for Computational Fluid Dynamics Application to Nuclear Reactor Thermal Hydraulics. NEA/CSNI/R(2016)4, OECD NEA Committee on the Safety of Nuclear Installations.
- CSNI** (2017a) Post-BEMUSE Reflood Model Input Uncertainty Methods (PREMIUM) Benchmark Final Report. NEA/CSNI/R(2016)18, OECD NEA Committee on the Safety of Nuclear Installations.
- CSNI** (2017b) A State-of-the-Art Report on Scaling in System Thermal-Hydraulics Applications to Nuclear Reactor Safety and Design. NEA/CSNI/R(2016)14, OECD NEA Committee on the Safety of Nuclear Installations.
- CSNI** (2018) Phenomena Identification and Ranking Table. R&D Priorities for Loss-of-Cooling and Loss-of-Coolant Accidents in Spent Nuclear Fuel Pools. NEA/CSNI/R(2017)18, OECD NEA Committee on the Safety of Nuclear Installations.
- CSNI** (2019) NEA Benchmark Exercise: Computational Fluid Dynamic Prediction and Uncertainty Quantification of a GEMIX Mixing Layer Test. NEA/CSNI/R(2017)19, OECD NEA Committee on the Safety of Nuclear Installations.
- Cutrono Rakhimov A, Visser D, Komen E** (2019) Uncertainty Quantification Method for CFD Applied to the Turbulent Mixing of Two Water Layers – II: Deterministic Sampling for Input Uncertainty. *Nuclear Engineering and Design*, 348, 146–158, dx.doi.org/10.1016/j.nucengdes.2019.04.016.
- Dakota Software Training** (2016) Dakota Overview. SAND2015-6794 TR, Sandia National Laboratories.
- Dalbey K, Eldred M S, Geraci G, Jakeman J D, Maupin K A, Monschke J A, Seidl D T, Swiler L P, et al.** (2020) Dakota A Multilevel Parallel Object-Oriented Framework for Design Optimization Parameter Estimation Uncertainty Quantification and Sensitivity Analysis: Version 6.12 Theory Manual. SAND2020-4987, Sandia National Laboratories.
- Darby J L** (2018) Applications of Evidence Theory to Issues with Nuclear Weapons. SAND2018-12985C, Sandia National Laboratories.
- D'Auria F** (2017) Thermal-Hydraulics of Water Cooled Nuclear Reactors. Woodhead.
- Debusschere B** (2017) Polynomial Chaos Based Uncertainty Quantification. SAND2017-8196C, Sandia National Laboratories.
- DeGroot M H, Schervish M J** (2012) Probability and Statistics. Pearson Modern Classic, fourth edition, Pearson Education, Inc.
- Deissler R G** (1992) Turbulent Fluid Motion IV-averages, Reynolds Decomposition, and the Closure Problem. NASA-TM-105822, National Aeronautics and Space Administration.
- Diamond D J, Brown N R, Denning R, Bajorek S** (2018) Phenomena Important in Modeling and Simulation of Molten Salt Reactors. BNL-114869-2018-IR, Brookhaven National Laboratory.
- Dolci V, Arina R** (2016) Proper Orthogonal Decomposition as Surrogate Model for Aerodynamic Optimization. *International Journal of Aerospace Engineering*, 2016, 1–15, dx.doi.org/10.1155/2016/8092824.

References

- Downing J, Jones J, Tehrani A, Moscrop R** (2018) A UK Regulatory Perspective on Computational Fluid Dynamics for Nuclear Safety Analysis. In *Proceedings of TOPFUEL 2018*.
- Duraisamy K, Iaccarino G, Xiao H** (2019) Turbulence Modeling in the Age of Data. *Annual Review of Fluid Mechanics*, 51(1), 357–377, [dx.doi.org/10.1146/annurev-fluid-010518-040547](https://doi.org/10.1146/annurev-fluid-010518-040547).
- Dzodzo M** (2019) Scaling Analysis and Relation to EMDAP and BEPU. *Nuclear Engineering and Design*, 353, 110228, [dx.doi.org/10.1016/j.nucengdes.2019.110228](https://doi.org/10.1016/j.nucengdes.2019.110228).
- Dzodzo M** (2021) Validation of Computational Fluid Dynamics Models for Industrial Applications. In *Proceedings of the 28th International Conference on Nuclear Engineering (ICONE 28)*.
- Efron B, Gong G** (1983) A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *The American Statistician*, 37(1), 36–48, [dx.doi.org/10.1080/00031305.1983.10483087](https://doi.org/10.1080/00031305.1983.10483087).
- Efron B, Tibshirani R** (1993) An Introduction to the Bootstrap. Chapman & Hall.
- EMS** (2021) Bhattacharyya Distance. European Mathematical Society Encyclopedia of Mathematics. https://encyclopediaofmath.org/wiki/Bhattacharyya_distance.
- Ferson S, Kreinovich V, Hajagos J, Oberkampf W, Ginzburg L** (2007) Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty. SAND2007-0939, Sandia National Laboratories, [dx.doi.org/10.2172/910198](https://doi.org/10.2172/910198).
- Frazer-Nash** (2019) Project FORTE - Nuclear Thermal Hydraulics Research and Development: Critical Review of State-of-the-Art Thermal Hydraulic Prediction Capability. FNC 53798/46733R Issue 1, Frazer-Nash Consultancy.
- Gelman A, Carlin J B, Stern H S, Dunson D B, Vehtari A, Rubin D B** (2014) Bayesian Data Analysis. Third edition, Chapman & Hall/CRC Press.
- Georgaka S, Stabile G, Star K, Rozza G, Bluck M J** (2020) A Hybrid Reduced Order Method for Modelling Turbulent Heat Transfer Problems. *Computers & Fluids*, 208, 104615, [dx.doi.org/10.1016/j.compfluid.2020.104615](https://doi.org/10.1016/j.compfluid.2020.104615).
- Gerstner T, Griebel M** (1998) Numerical Integration Using Sparse Grids. *Numerical Algorithms*, 18(3), 209, [dx.doi.org/10.1023/A:1019129717644](https://doi.org/10.1023/A:1019129717644).
- Giunta A, Wojtkiewicz S, Eldred M** (2003) Overview of Modern Design of Experiments Methods for Computational Simulations. In *41st Aerospace Sciences Meeting and Exhibit*, American Institute of Aeronautics and Astronautics, [dx.doi.org/10.2514/6.2003-649](https://doi.org/10.2514/6.2003-649).
- Giunta A A, McFarland J M, Swiler L P, Eldred M S** (2006) The Promise and Peril of Uncertainty Quantification Using Response Surface Approximations. *Structure and Infrastructure Engineering*, 2(3-4), 175–189, [dx.doi.org/10.1080/15732470600590507](https://doi.org/10.1080/15732470600590507).
- Glaeser H** (2008) GRS Method for Uncertainty and Sensitivity Evaluation of Code Results and Applications. *Science and Technology of Nuclear Installations*, 2008, 1–7, [dx.doi.org/10.1155/2008/798901](https://doi.org/10.1155/2008/798901).
- Glen G, Isaacs K** (2012) Estimating Sobol Sensitivity Indices Using Correlations. *Environmental Modelling & Software*, 37, 157–166, [dx.doi.org/10.1016/j.envsoft.2012.03.014](https://doi.org/10.1016/j.envsoft.2012.03.014).

References

- Görtler J, Kehlbeck R, Deussen O** (2019) A Visual Exploration of Gaussian Processes. *Distill*, 4(4), [dx.doi.org/10.23915/distill.00017](https://doi.org/10.23915/distill.00017).
- Gosling J P** (2018) SHELF: The Sheffield Elicitation Framework. In **Dias L C, Morton A, Quigley J** (editors) *Elicitation*, volume 261, pp. 61–93, Springer International Publishing, [dx.doi.org/10.1007/978-3-319-65052-4_4](https://doi.org/10.1007/978-3-319-65052-4_4).
- Hanna B N, Dinh N T, Youngblood R W, Bolotnov I A** (2020) Machine-Learning Based Error Prediction Approach for Coarse-Grid Computational Fluid Dynamics (CG-CFD). *Progress in Nuclear Energy*, 118, 103140, [dx.doi.org/10.1016/j.pnucene.2019.103140](https://doi.org/10.1016/j.pnucene.2019.103140).
- Helton J C** (2009) Conceptual and Computational Basis for the Quantification of Margins and Uncertainty. SAND2009-3055, Sandia National Laboratories.
- Hemming V, Burgman M A, Hanea A M, McBride M F, Wintle B C** (2018) A Practical Guide to Structured Expert Elicitation Using the IDEA Protocol. *Methods in Ecology and Evolution*, 9(1), 169–180, [dx.doi.org/10.1111/2041-210X.12857](https://doi.org/10.1111/2041-210X.12857).
- Hills R, Witkowski W, Urbina A, Rider W, Trucano T** (2013) Development of a Fourth Generation Predictive Capability Maturity Model. SAND2013-8051, Sandia National Laboratories, [dx.doi.org/10.2172/1096515](https://doi.org/10.2172/1096515).
- Hofer E** (1999) Sensitivity Analysis in the Context of Uncertainty Analysis for Computationally Intensive Models. *Computer Physics Communications*, 117(1-2), 21–34, [dx.doi.org/10.1016/S0010-4655\(98\)00153-2](https://doi.org/10.1016/S0010-4655(98)00153-2).
- Hofer E** (2018) The Uncertainty Analysis of Model Results. Springer International Publishing, [dx.doi.org/10.1007/978-3-319-76297-5](https://doi.org/10.1007/978-3-319-76297-5).
- Hosder S, Walters R W** (2007) Non-Intrusive Polynomial Chaos Methods for Stochastic CFD – Theory and Applications. In *Computational Uncertainty in Military Vehicle Design*.
- IAEA** (1991) Safety Culture. INSAG-4, International Atomic Energy Agency.
- IAEA** (2002) Accident Analysis for Nuclear Power Plants. Safety Reports Series No. 23, International Atomic Energy Agency.
- IAEA** (2003) Safety Margins of Operating Reactors. Analysis of Uncertainties and Implications for Decision Making. IAEA-TECDOC-1332, International Atomic Energy Agency.
- IAEA** (2008) Best Estimate Safety Analysis for Nuclear Power Plants: Uncertainty Evaluation. Safety Reports Series No. 52, International Atomic Energy Agency.
- IAEA** (2013) Benchmark Analyses on the Natural Circulation Test Performed during the PHENIX End-of-life Experiments. IAEA-TECDOC-1703, International Atomic Energy Agency.
- IAEA** (2014) Benchmark Analyses of Sodium Natural Convection in the Upper Plenum of the Monju Reactor Vessel. IAEA-TECDOC-1754, International Atomic Energy Agency.
- IAEA** (2017) Benchmark Analysis of EBR-II Shutdown Heat Removal Tests. IAEA-TECDOC-1819, International Atomic Energy Agency.
- IAEA** (2019) Deterministic Safety Analysis for Nuclear Power Plants. SSG-2 (Rev.1), International Atomic Energy Agency.

References

- IAEA** (2020a) Benchmarking of Computational Fluid Dynamics Codes for Fuel Assembly Design. IAEA-TECDOC-1907, International Atomic Energy Agency.
- IAEA** (2020b) Benchmarking of Computational Fluid Dynamics Codes for Reactor Vessel Design. IAEA-TECDOC-1908, International Atomic Energy Agency.
- IPCC** (2010) Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change.
- IPCC** (2014) Evaluation of Climate Models. In *Climate Change 2013 - The Physical Science Basis*, pp. 741–866, Intergovernmental Panel on Climate Change, [dx.doi.org/10.1017/CBO9781107415324.020](https://doi.org/10.1017/CBO9781107415324.020).
- Ivanov E, Bruna G, Sargeni A, Dubois F** (2019) Best Estimate Plus Uncertainty (BEPU): Why It Is Still Not Widely Used. *Nuclear Technology*, 205(12), 1529–1539, [dx.doi.org/10.1080/00295450.2019.1627122](https://doi.org/10.1080/00295450.2019.1627122).
- JCGM** (2008) Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement. JCGM 100:2008 (GUM 1995), Joint Committee for Guides in Metrology.
- Jiang W, Hales J D, Spencer B W, Collin B P, Slaughter A E, Novascone S R, Toptan A, Gamble K A, et al.** (2021) TRISO Particle Fuel Performance and Failure Analysis with BISON. *Journal of Nuclear Materials*, 548, 152795, [dx.doi.org/10.1016/j.jnucmat.2021.152795](https://doi.org/10.1016/j.jnucmat.2021.152795).
- Jones C, Hetzler A, Dinh N, Athe P, Sieger M** (2018) Updated Verification and Validation Assessment for VERA. SAND–2018-8800R, Consortium for Advanced Simulation of LWRs, [dx.doi.org/10.2172/1466526](https://doi.org/10.2172/1466526).
- Kahneman D** (2011) Thinking, Fast and Slow. Penguin Books.
- Keefer D L, Bodily S E** (1983) Three-Point Approximations for Continuous Random Variables. *Management Science*, 29(5), 595–609.
- Khalil I, Pratt Q, Schmachtenberger H, Ghanem R** (2018) Heat Transfer Modeling of Spent Nuclear Fuel Using Uncertainty Quantification and Polynomial Chaos Expansion. *Journal of Heat Transfer*, 140(2), 022001, [dx.doi.org/10.1115/1.4037501](https://doi.org/10.1115/1.4037501).
- Kiureghian A D, Ditlevsen O** (2009) Aleatory or Epistemic? Does It Matter? *Structural Safety*, 31(2), 105–112, [dx.doi.org/10.1016/j.strusafe.2008.06.020](https://doi.org/10.1016/j.strusafe.2008.06.020).
- Kreinovich V, Beck J, Ferregut C, Sanchez A, Keller G R, Averill M, Starks S A** (2006) Monte-Carlo-Type Techniques for Processing Interval Uncertainty, and Their Potential Engineering Applications. *Reliable Computing*, 13(1), 25–69, [dx.doi.org/10.1007/s11155-006-9021-6](https://doi.org/10.1007/s11155-006-9021-6).
- Krueger A M, Mousseau V A, Hassan Y A** (2019) Rigorous Code Verification: An Additional Tool to Use With the Method of Manufactured Solutions. In *ASME 2019 Verification and Validation Symposium*, American Society of Mechanical Engineers, [dx.doi.org/10.1115/VVS2019-5166](https://doi.org/10.1115/VVS2019-5166).
- Kruschke J K** (2015) Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. Second edition, Academic Press.
- Ku H H** (1966) Notes on the Use of Propagation of Error Formulas. *Journal of Research of rgw National Bureau of Standards-C. Engineering and Instrumentation*, 70C(4), 263–273.

References

- Kucherenko S, Iooss B** (2017) Derivative-Based Global Sensitivity Measures. In **Ghanem R, Higdon D, Owhadi H** (editors) *Handbook of Uncertainty Quantification*, pp. 1241–1263, Springer International Publishing, [dx.doi.org/10.1007/978-3-319-12385-1_36](https://doi.org/10.1007/978-3-319-12385-1_36).
- Lee E T, Wang J W** (2003) Statistical Methods for Survival Data Analysis. Wiley Series in Probability and Statistics, third edition, John Wiley & Sons, Inc., [dx.doi.org/10.1002/0471458546](https://doi.org/10.1002/0471458546).
- Li J, Liu Y Y** (2016) Thermal Modeling of a Vertical Dry Storage Cask for Used Nuclear Fuel. *Nuclear Engineering and Design*, 301, 74–88, [dx.doi.org/10.1016/j.nucengdes.2016.01.008](https://doi.org/10.1016/j.nucengdes.2016.01.008).
- Lin H C, Zhang S, Diamond D, Bajorek S, Christensen R, Guo Y, Yoder G, Shi S, et al.** (2019) Phenomena Identification and Ranking Table Study for Thermal Hydraulics for Advanced High Temperature Reactor. *Annals of Nuclear Energy*, 124, 257–269, [dx.doi.org/10.1016/j.anucene.2018.08.038](https://doi.org/10.1016/j.anucene.2018.08.038).
- Lin L, Dinh N** (2020) Predictive Capability Maturity Quantification Using Bayesian Network. *Journal of Verification, Validation and Uncertainty Quantification*, 5(3), 031001, [dx.doi.org/10.1115/1.4048465](https://doi.org/10.1115/1.4048465).
- MacKay D J C** (2003) Information Theory, Inference, and Learning Algorithms. Cambridge University Press.
- Mahadevan S, Sarkar S** (2009) Uncertainty Analysis Methods. CBP-TR-2009-002, Rev. 0 attachment to IAEA-TECDOC-1701, Vanderbilt University.
- Makarashvili V, Merzari E, Obabko A, Siegel A, Fischer P** (2017) A Performance Analysis of Ensemble Averaging for High Fidelity Turbulence Simulations at the Strong Scaling Limit. *Computer Physics Communications*, 219, 236–245, [dx.doi.org/10.1016/j.cpc.2017.05.023](https://doi.org/10.1016/j.cpc.2017.05.023).
- Martin R P** (2011) Quantifying Phenomenological Importance in Best-Estimate Plus Uncertainty Analyses. *Nuclear Technology*, 175(3), 652–662, [dx.doi.org/10.13182/NT175-652](https://doi.org/10.13182/NT175-652).
- Martin R P, Frepoli C** (2019) Design-Basis Accident Analysis Methods for Light-Water Nuclear Power Plants. World Scientific, [dx.doi.org/10.1142/11139](https://doi.org/10.1142/11139).
- Merzari E, Fischer P, Min M, Kerkemeier S, Obabko A, Shaver D, Yuan H, Yu Y, et al.** (2020) Toward Exascale: Overview of Large Eddy Simulations and Direct Numerical Simulations of Nuclear Reactor Flows with the Spectral Element Method in Nek5000. *Nuclear Technology*, 206(9), 1308–1324, [dx.doi.org/10.1080/00295450.2020.1748557](https://doi.org/10.1080/00295450.2020.1748557).
- Merzari E, Ninokata H** (2011) Proper Orthogonal Decomposition of the Flow in a Tight Lattice Rod-Bundle. *Nuclear Engineering and Design*, 241(11), 4621–4632, [dx.doi.org/10.1016/j.nucengdes.2010.12.005](https://doi.org/10.1016/j.nucengdes.2010.12.005).
- Merzari E, Pointer W D, Fischer P** (2013) Numerical Simulation and Proper Orthogonal Decomposition of the Flow in a Counter-Flow T-Junction. *Journal of Fluids Engineering*, 135(9), 091304, [dx.doi.org/10.1115/1.4024059](https://doi.org/10.1115/1.4024059).
- Miller D S** (2009) Internal Flow Systems. Second edition, Mentor Graphics.
- Moore R E, Kearfott R B, Cloud M J** (2009) Introduction to Interval Analysis. Society for Industrial and Applied Mathematics, [dx.doi.org/10.1137/1.9780898717716](https://doi.org/10.1137/1.9780898717716).

References

- Morgan M G** (2014) Use (and Abuse) of Expert Elicitation in Support of Decision Making for Public Policy. *Proceedings of the National Academy of Sciences*, 111(20), 7176–7184, dx.doi.org/10.1073/pnas.1319946111.
- Morris D E, Oakley J E, Crowe J A** (2014) A Web-Based Tool for Eliciting Probability Distributions from Experts. *Environmental Modelling & Software*, 52, 1–4, dx.doi.org/10.1016/j.envsoft.2013.10.010.
- Mui T, Hu R, Zhang G** (2019) Uncertainty Quantification on SAM Simulations of EBR-II Loss-of-Flow Tests. In *Proceedings of 18th International Topical Meeting on Nuclear Reactor Thermal Hydraulics (NURETH-18)*.
- Najm H N** (2009) Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics. *Annual Review of Fluid Mechanics*, 41(1), 35–52, dx.doi.org/10.1146/annurev.fluid.010908.165248.
- NASA** (2016) Standard for Models and Simulations. NASA-STD-7009A W/CHANGE 1, National Aeronautics and Space Administration.
- Navidi W C** (2020) Statistics for Engineers and Scientists. Fifth edition, McGraw-Hill Education.
- Nguyen T, Vaghetto R, Hassan Y** (2020) Experimental Investigation of Turbulent Wake Flows in a Helically Wrapped Rod Bundle in Presence of Localized Blockages. *Physics of Fluids*, 32(7), 075113, dx.doi.org/10.1063/5.0008589.
- NIST** (2021) E-Handbook of Statistical Methods. NIST/SEMATECH, dx.doi.org/10.18434/M32189.
- Nouy E, de Crécy A** (2017) Quantification of the Uncertainty of Physical Models Integrated into System Thermohydraulic Codes. *Nuclear Engineering and Design*, 321, 278–287, dx.doi.org/10.1016/j.nucengdes.2016.10.032.
- Oberkampf W L** (2002) Discussion: “Comprehensive Approach to Verification and Validation of CFD Simulations—Part 1: Methodology and Procedures” (Stern, F., Wilson, R. V., Coleman, H. W., and Paterson, E. G., 2001, ASME J. Fluids Eng., 123, 793–802). *Journal of Fluids Engineering*, 124(3), 809–810, dx.doi.org/10.1115/1.1492826.
- Oberkampf W L, Barone M F** (2006) Measures of Agreement between Computation and Experiment: Validation Metrics. *Journal of Computational Physics*, 217(1), 5–36, dx.doi.org/10.1016/j.jcp.2006.03.037.
- Oberkampf W L, Roy C J** (2010) Verification and Validation in Scientific Computing. Cambridge University Press, dx.doi.org/10.1017/CBO9780511760396.
- Oberkampf W L, Trucano T G** (2002) Verification and Validation in Computational Fluid Dynamics. SAND2002-0529, Sandia National Laboratories, dx.doi.org/10.2172/793406.
- Oberkampf W L, Trucano T G** (2003) Verification, Validation, and Predictive Capability in Computational Engineering and Physics. SAND2003-3769, Sandia National Laboratories.
- Oberkampf W L, Trucano T G** (2007) Verification and Validation Benchmarks. SAND2007-0853, Sandia National Laboratories.

References

- Oberkampff W L, Trucano T G, Pilch M M** (2007) Predictive Capability Maturity Model for Computational Modeling and Simulation. SAND2007-5948, Sandia National Laboratories, dx.doi.org/10.2172/976951.
- ONR** (2019) Validation of Computer Codes and Calculation Methods. NS-TAST-GD-042 Revision 4, Office for Nuclear Regulation.
- Orea D, Vaghetto R, Nguyen T, Hassan Y** (2020) Experimental Measurements of Flow Mixing in Cold Leg of a Pressurized Water Reactor. *Annals of Nuclear Energy*, 140, 107137, dx.doi.org/10.1016/j.anucene.2019.107137.
- ORNL** (2008a) Next Generation Nuclear Plant Phenomena Identification and Ranking Tables (PIRTs) Volume 1: Main Report. ORNL/TM-2007/147 Vol 1, Oak Ridge National Laboratory, dx.doi.org/10.2172/1001275.
- ORNL** (2008b) Next Generation Nuclear Plant Phenomena Identification and Ranking Tables (PIRTs) Volume 2: Accident and Thermal Fluids Analysis PIRTs. ORNL/TM-2007/147 Vol 2, Oak Ridge National Laboratory, dx.doi.org/10.2172/1001274.
- Pilch M M, Trucano T G, Helton J C** (2006) Ideas Underlying Quantification of Margins and Uncertainties (QMU). SAND2006-5001, Sandia National Laboratories, dx.doi.org/10.2172/891369.
- Porter N** (2019) Wilks' Formula Applied to Computational Tools: A Practical Discussion and Verification. *Annals of Nuclear Energy*, 133, 129–137, dx.doi.org/10.1016/j.anucene.2019.05.012.
- Pourgol-Mohamad M, Mosleh A, Modarres M** (2011) Structured Treatment of Model Uncertainty in Complex Thermal-Hydraulics Codes: Technical Challenges, Prospective and Characterization. *Nuclear Engineering and Design*, 241(1), 285–295, dx.doi.org/10.1016/j.nucengdes.2010.10.035.
- Rasmussen C E, Williams C K I** (2006) Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning, MIT Press.
- Ren W** (2013) Gen IV Materials Handbook Functionalities and Operation (4A) Handbook Version 4.0. ORNL/TM-2013/406, Oak Ridge National Laboratory, dx.doi.org/10.2172/1110968.
- Ren W** (2016) FY2017 Pilot Project Plan for the Nuclear Energy Knowledge and Validation Center Initiative. ORNL/TM-2016/615, Oak Ridge National Laboratory, dx.doi.org/10.2172/1333667.
- Rider B, Hills R, Trucano T, Ubrina A, Witkowski W** (2013) Improving Flexibility and Applicability of Modeling and Simulation Credibility Assessments (via PCMM). SAND 2013-0307A, Sandia National Laboratories.
- Roache P J** (1997) Quantification of Uncertainty in Computational Fluid Dynamics. *Annual Review of Fluid Mechanics*, 29(1), 123–160, dx.doi.org/10.1146/annurev.fluid.29.1.123.
- Roache P J** (2016) Verification and Validation in Fluids Engineering: Some Current Issues. *Journal of Fluids Engineering*, 138(10), 101205, dx.doi.org/10.1115/1.4033979.
- Roelofs F** (2019) Thermal Hydraulics Aspects of Liquid Metal Cooled Nuclear Reactors. Woodhead.
- Ross T J** (2010) Fuzzy Logic with Engineering Applications. Third edition, John Wiley.

References

- Roy C J** (2005) Review of Code and Solution Verification Procedures for Computational Simulation. *Journal of Computational Physics*, 205(1), 131–156, dx.doi.org/10.1016/j.jcp.2004.10.036.
- Roy C J, Oberkampf W L** (2011) A Comprehensive Framework for Verification, Validation, and Uncertainty Quantification in Scientific Computing. *Computer Methods in Applied Mechanics and Engineering*, 200(25-28), 2131–2144, dx.doi.org/10.1016/j.cma.2011.03.016.
- Runchal A** (editor) (2020) 50 Years of CFD in Engineering Sciences: A Commemorative Volume in Memory of D. Brian Spalding. Springer, dx.doi.org/10.1007/978-981-15-2670-1.
- Salari K, Knupp P** (2000) Code Verification by the Method of Manufactured Solutions. SAND2000-1444, Sandia National Laboratories, dx.doi.org/10.2172/759450.
- Salko R K, Blyth T S, Dances C A, Magedanz J W, Jernigan C, Kelly J, Toptan A, Gergar M, et al.** (2016) CTF Validation and Verification. CASL-U-2016-1113-000, U.S. Department of Energy.
- Sallaberry C, Helton J, Hora S** (2008) Extension of Latin Hypercube Samples with Correlated Variables. *Reliability Engineering & System Safety*, 93(7), 1047–1059, dx.doi.org/10.1016/j.ress.2007.04.005.
- Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S** (2007) Global Sensitivity Analysis. the Primer. John Wiley & Sons, Ltd, dx.doi.org/10.1002/9780470725184.
- Schaefer J, Hosder S, West T, Rumsey C, Carlson J R, Kleb W** (2016) Uncertainty Quantification of Turbulence Model Closure Coefficients for Transonic Wall-Bounded Flows. *AIAA Journal*, 55(1), 195–213, dx.doi.org/10.2514/1.J054902.
- Schöbi R, Sudret B** (2017) Uncertainty Propagation of P-Boxes Using Sparse Polynomial Chaos Expansions. *Journal of Computational Physics*, 339, 307–327, dx.doi.org/10.1016/j.jcp.2017.03.021.
- Shafer G** (1976) A Mathematical Theory of Evidence. Princeton University Press.
- Shields M D, Zhang J** (2016) The Generalization of Latin Hypercube Sampling. *Reliability Engineering & System Safety*, 148, 96–108, dx.doi.org/10.1016/j.ress.2015.12.002.
- Singh P M, Chan K J, Deo C S, Deodeshmukh V, Keiser J R, Ren W, Sham T, Wilson D F, et al.** (2019) Phenomena Identification and Ranking Table (PIRT) Study for Metallic Structural Materials for Advanced High-Temperature Reactor. *Annals of Nuclear Energy*, 123, 222–229, dx.doi.org/10.1016/j.anucene.2018.08.036.
- Skorek T, de Crécy A, Kovtonyuk A, Petruzzi A, Mendizábal R, de Alfonso E, Reventós F, Freixa J, et al.** (2019) Quantification of the Uncertainty of the Physical Models in the System Thermal-Hydraulic Codes – PREMIUM Benchmark. *Nuclear Engineering and Design*, 354, 110199, dx.doi.org/10.1016/j.nucengdes.2019.110199.
- Smith J E** (1993) Moment Methods for Decision Analysis. *Management Science*, 39(3), 340–358, dx.doi.org/10.1287/mnsc.39.3.340.
- Stern F, Wilson R V, Coleman H W, Paterson E G** (2001) Comprehensive Approach to Verification

References

- and Validation of CFD Simulations—Part 1: Methodology and Procedures. *Journal of Fluids Engineering*, 123(4), 793–802, dx.doi.org/10.1115/1.1412235.
- Sudret B** (2008) Global Sensitivity Analysis Using Polynomial Chaos Expansions. *Reliability Engineering & System Safety*, 93(7), 964–979, dx.doi.org/10.1016/j.ress.2007.04.002.
- Swiler L P, Giunta A A** (2007) Aleatory and Epistemic Uncertainty Quantification for Engineering Applications. SAND2007-4655C, Sandia National Laboratories.
- Tennøe S, Haldnes G, Einevoll G T** (2018) Uncertainpy: A Python Toolbox for Uncertainty Quantification and Sensitivity Analysis in Computational Neuroscience. *Frontiers in Neuroinformatics*, 12, 49, dx.doi.org/10.3389/fninf.2018.00049.
- Tucker W T, Ferson S** (2003) Probability Bounds Analysis in Environmental Risk Assessments. Applied Biomathematics.
- Tucker W T, Ferson S** (2006) Sensitivity in Risk Analyses with Uncertain Numbers. SAND2006-2801, Sandia National Laboratories, dx.doi.org/10.2172/886899.
- Uchida K, Kawamura S** (2008) Error Estimation of Values Obtained from Practical CFD Analysis. *Journal of Nuclear Science and Technology*, 45(7), 625–633, dx.doi.org/10.3327/jnst.45.625.
- US NRC** (1989) Quantifying Reactor Safety Margins Application of Code Scaling, Applicability, and Uncertainty Evaluation Methodology to a Large-Break, Loss-of-Coolant Accident. NUREG/CR-5249 Rev. No. 4, U.S. Nuclear Regulatory Commission.
- US NRC** (2004) TRISO-Coated Particle Fuel Phenomenon Identification and Ranking Tables (PIRTs) for Fission Product Transport Due to Manufacturing, Operations, and Accidents: Main Report. NUREG/CR-6844, Volume 1, U.S. Nuclear Regulatory Commission.
- US NRC** (2005) Transient and Accident Analysis Methods. Regulatory Guide 1.203, U.S. Nuclear Regulatory Commission.
- US NRC** (2013) Computational Fluid Dynamics Best Practice Guidelines for Dry Cask Applications. NUREG-2152, U.S. Nuclear Regulatory Commission.
- US NRC** (2019a) CFD Validation of Vertical Dry Cask Storage System. NUREG/CR-7260, U.S. Nuclear Regulatory Commission.
- US NRC** (2019b) Credibility Assessment Framework for Critical Boiling Transition Models: A Generic Safety Case to Determine the Credibility of Critical Heat Flux and Critical Power Models. NUREG/KM-0013 Draft for Comment, U.S. Nuclear Regulatory Commission.
- US NRC** (2020) NRC Non-Light Water Reactor (Non-LWR) Vision and Strategy, Volume 1 – Computer Code Suite for Non-LWR Design Basis Event Analysis. Revision 1, U.S. Nuclear Regulatory Commission.
- Vose D** (2008) Risk Analysis: A Quantitative Guide. Third edition, Wiley.
- Walters R W, Huyse L** (2002) Uncertainty Analysis for Fluid Mechanics with Applications. NASA/CR-2002-211449, ICASE Report No. 2002-1, National Aeronautics and Space Administration.

References

- Waskom M** (2021) Seaborn: Statistical Data Visualization. *Journal of Open Source Software*, 6(60), 3021, [dx.doi.org/10.21105/joss.03021](https://doi.org/10.21105/joss.03021).
- Weirs G, Winokur J** (2017) A Summary of Modeling and Simulation Assessment Techniques. SAND2017-7023PE, Sandia National Laboratories.
- Weirs V G, Mousseau K C, Johnson R W, Lee H, Oberkampf W L, Smith B, Fort J, Rider W J** (2011) Nuclear Energy – Knowledge Base for Advanced Modeling and Simulation (NE-KAMS) Code Verification and Validation Data Standards and Requirements: Fluid Dynamics Version 1.0. INL/EXT-11-23419, Idaho National Laboratory, [dx.doi.org/10.2172/1033905](https://doi.org/10.2172/1033905).
- Weiss J** (2019) A Tutorial on the Proper Orthogonal Decomposition. In *AIAA Aviation 2019 Forum*, American Institute of Aeronautics and Astronautics, [dx.doi.org/10.2514/6.2019-3333](https://doi.org/10.2514/6.2019-3333).
- Wenig P J, Ji R, Kelm S, Klein M** (2021) Towards Uncertainty Quantification of LES and URANS for the Buoyancy-Driven Mixing Process between Two Miscible Fluids—Differentially Heated Cavity of Aspect Ratio 4. *Fluids*, 6(4), 161, [dx.doi.org/10.3390/fluids6040161](https://doi.org/10.3390/fluids6040161).
- Wilks S S** (1942) Statistical Prediction with Special Reference to the Problem of Tolerance Limits. *The Annals of Mathematical Statistics*, 13(4), 400–409, [dx.doi.org/10.1214/aoms/1177731537](https://doi.org/10.1214/aoms/1177731537).
- Wilson G E** (2013) Historical Insights in the Development of Best Estimate Plus Uncertainty Safety Analysis. *Annals of Nuclear Energy*, 52, 2–9, [dx.doi.org/10.1016/j.anucene.2012.03.002](https://doi.org/10.1016/j.anucene.2012.03.002).
- Wilson G E, Boyack B E** (1998) The Role of the PIRT Process in Experiments, Code Development and Code Applications Associated with Reactor Safety Analysis. *Nuclear Engineering and Design*, 186(1), 23–37, [dx.doi.org/10.1016/S0029-5493\(98\)00216-7](https://doi.org/10.1016/S0029-5493(98)00216-7).
- Wright R F, Swantner S, Swartz M M, Lojek J, Song Y J, Kindred T A** (2014) AP1000® Passive Core Cooling System Performance Following an Extended Station Blackout Event. In *Volume 3: Next Generation Reactors and Advanced Reactors; Nuclear Safety and Security*, p. V003T06A050, American Society of Mechanical Engineers, [dx.doi.org/10.1115/ICONE22-311129](https://doi.org/10.1115/ICONE22-311129).
- Wu X, Kozlowski T, Meidani H, Shirvan K** (2018) Inverse Uncertainty Quantification Using the Modular Bayesian Approach Based on Gaussian Process, Part 1: Theory. *Nuclear Engineering and Design*, 335, 339–355, [dx.doi.org/10.1016/j.nucengdes.2018.06.004](https://doi.org/10.1016/j.nucengdes.2018.06.004).
- Yeh S T, Wang X, Sung C L, Mak S, Chang Y H, Zhang L, Wu C F J, Yang V** (2018) Common Proper Orthogonal Decomposition-Based Spatiotemporal Emulator for Design Exploration. *AIAA Journal*, 56(6), 2429–2442, [dx.doi.org/10.2514/1.J056640](https://doi.org/10.2514/1.J056640).
- Yurko J P, Buongiorno J** (2012) Quantitative Phenomena Identification and Ranking Table (QPIRT) for Bayesian Uncertainty Quantification. In *Proceedings of the 2012 International Congress on Advances in Nuclear Power Plants (ICAPP'12)*, American Nuclear Society.

6 Abbreviations

AIC	Akaike Information Criterion
ASME	American Society of Mechanical Engineers
BEMUSE	Best Estimate Methods Uncertainty and Sensitivity Evaluation
BEPU	Best Estimate Plus Uncertainty
BWR	Boiling Water Reactor
C&I	Control & Instrumentation
CAD	Computer Aided Design
CASL	Consortium for Advanced Simulation of Light Water Reactors
CCDF	Complementary Cumulative Distribution Function
CDF	Cumulative Distribution Function
CFD	Computational Fluid Dynamics
CFL	Courant-Friedrichs Lewy
CHT	Conjugate Heat Transfer
CI	Confidence Interval
CLT	Central Limit Theorem
CPU	Central Processing Unit
CSAU	Code Scaling, Applicability and Uncertainty
CSNI	Committee on the Safety of Nuclear Installations
DACE	Design and Analysis of Computer Experiments
DNS	Direct Numerical Simulation
DoE	Design of Experiments
DST	Dempster-Shafer Theory
EBR	Experimental Breeder Reactor
EMDAP	Evaluation Methodology Development and Application Process
FEA	Finite Element Analysis
FORM	First Order Reliability Method
FSI	Fluid-Structure Interaction
GCI	Grid Convergence Index
Gen IV	Generation IV
GRS	Gesellschaft für Anlagen- und Reaktorsicherheit
IAEA	International Atomic Energy Agency
IET	Integral Effect Test
IUQ	Inverse Uncertainty Quantification
IVP	Interval-Valued Probability
KDE	Kernel Density Estimation
LES	Large Eddy Simulation
LHS	Latin Hypercube Sampling
LOCA	Loss-Of-Coolant Accident
LWR	Light Water Reactor

Abbreviations

M&S	Modelling and Simulation
ML	Machine Learning
MLE	Maximum Likelihood Estimation
MMS	Method of Manufactured Solutions
MSR	Molten Salt Reactor
NEA	Nuclear Energy Agency
NGNP	Next Generation Nuclear plant
NPP	Nuclear Power Plant
NTH	Nuclear Thermal Hydraulics
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares Regression
ONR	Office for Nuclear Regulation
PBA	Probability Bounds Analysis
PCE	Polynomial Chaos Expansion
PCMM	Predictive Capability Maturity Model
PCT	Peak Cladding Temperature
PDE	Partial Differential Equation
PDF	Probability Density Function
PIRT	Phenomena Identification and Ranking Table
PMF	Probability Mass Function
POD	Proper Orthogonal Decomposition
PREMIUM	Post-BEMUSE Reflood Model Input Uncertainty Methods
PSA	Probabilistic Safety Analysis
PWR	Pressurised Water Reactor
QA	Quality Assurance
QMU	Quantification of Margins and Uncertainties
RANS	Reynolds-Averaged Navier-Stokes
RMS	Root Mean Square
ROM	Reduced Order Model
SA	Sensitivity Analysis
SAPIUM	Systematic Approach for Input Uncertainty quantification Methodology
SET	Separate Effect Test
SFR	Sodium-cooled Fast Reactor
SORM	Second Order Reliability Method
SQA	Software Quality Assurance
SQE	Software Quality Engineering
SRQ	System Response Quantity
TRISO	TRIsstructural-ISOtropic
UMAE	Uncertainty Methodology based on Accuracy Extrapolation
UMS	Uncertainty Method Study
UQ	Uncertainty Quantification
URANS	Unsteady Reynolds-Averaged Navier-Stokes
US NRC	United States Nuclear Regulatory Commission
V&V	Verification and Validation

Abbreviations

VHTR Very High Temperature Reactor

